ETG-ETL Portfolio Optimization

Yun Zhang

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:

James V. Burke, Chair

R Douglas Martin

Ka Kit Tung

Hong Qian

Program Authorized to Offer Degree:
Applied Mathematics

University of Washington

Abstract

ETG-ETL Portfolio Optimization

Yun Zhang

Chair of the Supervisory Committee:
Professor James V. Burke
Department of Mathematics

Modern Portfolio Theory dates back to 1950s, when Markowitz proposed mean-variance portfolio optimization to construct portfolios. It provided a systematic approach to determine portfolio allocation when one is facing complicated risk structure that not only exists for individual assets but also across different assets. Since then there has been much research exploring better ways to quantify risk. In particular, asymmetric risk measures including the more recent downside risk measures. Here we use expected tail loss (ETL) as the risk measure which is a coherent risk measure, and define a reward measure, expected tail gain (ETG), to measure the upside return. We formulate the portfolio optimization problem using these two measures and developed an iterative algorithm to find its optimal solution.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to University of Washington, where I have had the opportunity to work with Professor James V. Burke from the Department of Mathematics, Professor Douglas Martin from the Department of Statistics, Professor Ka-Kit Tung and Professor Hong Qian from the Department of Applied Mathematics. Without their support, help and guidance, the accomplishment of the research represented in this thesis would not have been possible.

Chapter 1

INTRODUCTION

A portfolio is a combination of individual investment assets. It has two primary characteristics: reward and risk. The higher the reward is, the higher the risk, and vice versa. Portfolio optimization solves the problem of allocating weight to each member asset such that the portfolio's reward is maximized while its risk is less than some given value, or the portfolio's risk is minimized while its reward is greater than or equal to some given value. This general approach is referred to as Modern Portfolio Theory (MPT), originated by Markowitz who won the Nobel Prize for his pioneering work in this area. MPT provides a systematic way to determine portfolio allocation when one is facing complicated risk structure that not only exists for individual asset but also across different assets.

There has been much research exploring better ways to quantify risk since the establishment of MPT. From the earliest symmetric risk measures to later asymmetric risk that focus on the down-side risk. Here we use Expected Tail Loss (ETL), or Conditional Value-at-Risk (CVaR), as the risk measure, and defined a reward measure, Expected Tail Gain (ETG), in a manner similar to ETL. The resulting portfolio optimization problem turns out to be a concave programming problem. This problem can be converted into a mixed integer optimization problem, but in this form the problem requires an exhaustive search over all possible combinatorial options and is thus prohibitively time consuming even for small portfolios. We suggest a heuristic algorithm that is time efficient and provides excellent results based on empirical data for portfolios containing hundreds of stocks.

Chapter 2

RISK, REWARD AND PERFORMANCE MEASURES

2.1 Risk Measures

In finance, risk is the uncertainty that an investment's actual return will be different from expected. There are different types of risk associated with different financial assets, such as interest rate risk, credit risk, liquidity risk, currency risk etc. This thesis specifically focuses on equity risk, i.e., the uncertainty of future prices for equity securities. For a single asset, let P_t denote its price at t, and its *simple return* is defined as

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}, \tag{2.1.1}$$

which is also called *holding period return*. Its *log return* is defined as

$$r_t = \log \frac{P_t}{P_{t-1}}. \tag{2.1.2}$$

These formulas assume the price has been dividend and split adjusted. Notice that simple return has a minimum bound of -1. This is hard to handle when we use parameterized distributions, such as normal distribution or skewed t distribution, to model the return. The log return doesn't have this issue as its value can spread over the entire real axis. Therefore, the log return is preferred. Since r_t is a random variable given the price at P_{t-1}, risk can be defined as r_t's uncertainty. Throughout this thesis, we use r as the random return, $f(r)$ as its probability density function and $F(r)$ as its cumulative density function in the formulas for various risk measures.

This section lists a few concepts that are essential in understanding risk measures, including symmetric and asymmetric/downside risk measures. In particular, we focus on *coherent risk measures* of risk.

2.1.1 Classical Risk Measures

Volatility

Volatility is defined as the standard deviation of an asset's return. It is the most widely used measure of uncertainty and is usually denoted by σ,

$$\sigma = \sqrt{E(r - Er)^2},$$

where E stands for expectation. Volatility has the same units as the return and is a symmetric risk measure in the sense that it treats both the upside and downside deviation from the mean identically.

Mean Absolute Deviation (MAD)

Mean-absolute deviation is another symmetric measure, defined as the expected absolute deviation from the mean,

$$\text{MAD} \; = E|r - Er|.$$

It serves as a substitute for volitility in measuring the statistical dispersion when volitility is infinite. However, it is not everywhere differential and so is not preferred as an objective function in portfolio optimization.

Volatility and MAD are both symmetric risk measures, i.e., they treat positive and negative deviations from the mean the same. However, historical returns indicate that return is not normally distributed, but rather fat-tailed and skewed. In addition, investors favor a scenario where realized return is higher than expected, and consider as risk only a downside return that is less than a certain expected return. In this regard, numerous downside risk measures have been proposed.

Semi-standard Deviation (SSD)

Semi-standard deviation is similar to standard deviation except that it only takes into account the negative deviations from the mean,

$$\text{SSD} = \sqrt{\int_{-\infty}^{\mu} (r-\mu)^2 f(r)dr}.$$

Here μ is the expectation of r. Semi-standard Deviation was first proposed by Markowitz [25]. Quirk and Saposnik[33] provided a theoretical background for SSD being superior than variance/volitility. Mao [23] advocates using SSD in adapting Markowitz's model to capital budgeting.

Lower Partial Moments (LPM)

A key development in the research of downside risk measures was Lower Partial Moments (LPM) by Bawa[4] and Fishburn[12].

$$\text{LPM}(\alpha, c) = \int_{-\infty}^{c} (c-r)^{\alpha} f(r)dr \quad \alpha > 0$$

where c is the minimum acceptable return, a predetermined threshold, and α is the degree of the lower partial moment. We can set c to be a constant, say 0, or the risk-free rate, and α can take non-integer values. The LPM with $\alpha < 1$, $\alpha = 1$ and $\alpha > 1$ captures risk seeking behavior, risk neutral behavior and risk averse behavior, respectively. When $c = \mu$, the expectation of return r, and $\alpha = 2$, then $\text{LPM}(2, \mu)$ is just SSD.

Value-at-Risk (VaR)

Financial engineers at J.P.Morgan developed Value-at-Risk. It is widely used by practitioners to measure the risk of loss on a portfolio of financial assets. Given a time horizon and a probability level of α (usually 1% or 5%), $\text{VaR}_r(\alpha)$ is defined as the negative of the α-quantile,

$$\text{VaR}_{\alpha}(r) \triangleq -F^{-1}(\alpha) = -\inf\{x | F(x) \geq \alpha\}.$$

The value $\text{VaR}_r(\alpha)$ is interpreted as maximum loss in dollars for each dollar invested at $1-\alpha$ confidence level.

Expected Tail Loss (ETL)

Expected Tail Loss (ETL), introduced by Rockafellar and Uryasev [37, 38], is an alternative to VaR. Given a time horizon and a probability level α (usually 1% or 5%), $\text{ETL}_\alpha(r)$ is defined as the negative of the conditional expectation of the return r given that r is below the α-quantile, i.e. the negative of $\text{VaR}_\alpha(r)$,

$$\text{ETL}_\alpha(r) \triangleq -E[r|r \leq -\text{VaR}_\alpha(r)]$$

ETL is also called the Conditional Value-at-Risk(CVaR), the Expected Shortfall (ES), or the average Value-at-Risk. Figure 2.1 shows a illustration of VaR and ETL. ETL captures the tail returns beyond the VaR value, and thus captures extreme losses that VaR doesn't. ETL is always greater than or equal to VaR, and a large VaR indicates a large ETL. We prefer ETL to VaR due to its numerous favorable properties. In particular, because ETL is a *coherent* measure of risk.

2.1.2 Coherent Risk Measures

Artzner et al [29, 30] proposed four desirable properties for measures of risk and set up a framework for the study of risk measures.

Definition 2.1.1. For a probability space $(\Omega, \mathcal{F}, P)$, a *coherent risk measure* is a mapping $\rho : \chi \mapsto \bar{\mathbb{R}}$ satisfying the following four properties, where χ is the linear space of $\mathcal{F}$-measurable functions $X : \Omega \mapsto \mathbb{R}$.

A1. Translation invariance: $\rho(X+\alpha) = \rho(X) - \alpha$, for all $X \in \chi$ and all $\alpha \in \mathbb{R}$;

A2. Subadditivity: $\rho(X+Y) \leq \rho(X) + \rho(Y)$, for all X and $Y \in \chi$;

A3. Positive homogeneity: $\rho(\lambda X) = \lambda\rho(X)$, for all $\lambda \geq 0$ and all $X \in \chi$;

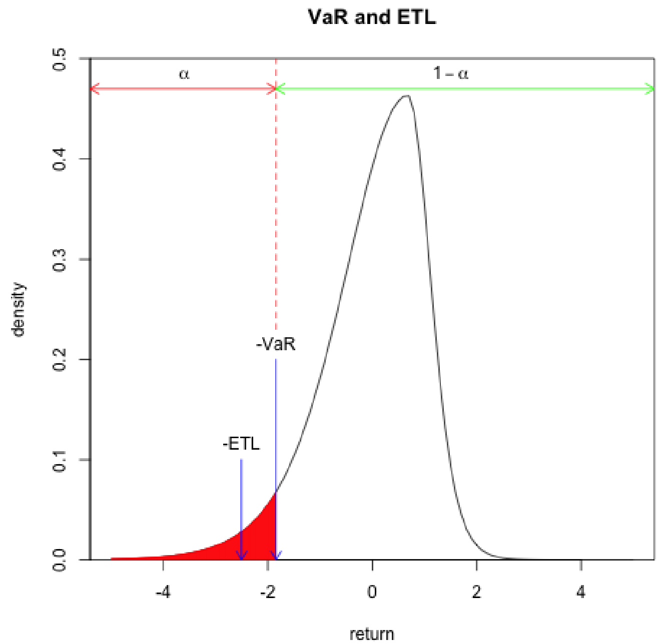

Figure 2.1: Illustration of VaR and ETL.

A4. Monotonicity: $\rho(X) \geq \rho(Y)$, for all X and $Y \in \chi$ with $X \leq Y$.

Here X, Y are taken as random variables of the return with $X \leq Y$ assumed to hold almost surely.

Remarks.

Criteria A1 means that adding a cash amount of α to the initial holdings reduces the risk measure by α.

Criteria A2 is a natural requirement that the risk of sum of two assets can only be less than or equal to the sum of individual risks, i.e., diversification reduces risk.

CriterionA3 implies that risk scales proportionally to position, i.e., when the initial holding amount is scaled, risk is scaled by exactly the same constant.

Critrion A4 indicates that greater loss implies larger risk.

Notice that A2 and A3 imply the convexity of ρ. The risk measure VaR does not satisfy subadditivity and so it is not convex and is not coherent. On the other hand, ETL satisfies all the four properties and so it is coherent [31]. For this reason it is preferred as measure of risk.

2.2 Reward Measures

Although there have been a great deal of research and discussion of risk measures in the literature, measures of reward have received considerably less attention. The consensus seems to be that the expected return $\mu = E(r)$ is a a sufficiently good measure of the reward. However, empirically, this measure is unstable on a number of data sets [27].

Given a series of historical returns of an asset $r_1, \ldots, r_T$, an unbiased estimate of the their expected return is the mean,

$$\hat{\mu} = \frac{1}{T} \sum_{i=1}^{T} r_i.$$

However, the sample mean is sensitive to outliers and thus is not a robust estimator. In this regard, we suggest using an asymmetric upside reward measure. For each downside risk measure discussed in previous section, a corresponding reward measure can be defined. The reward measure corresponding to SSD is

$$\sqrt{\int_{\mu}^{\infty} (r-\mu)^2 f(r) dr},$$

the reward measure corresponding to LPM is

$$\int_{c}^{\infty} (c-r)^{\alpha} f(r) dr \quad \alpha > 0 \ ,$$

and the reward measure corresponding to VaR is

$$\inf\{x | F(x) \geq 1-\beta\},$$

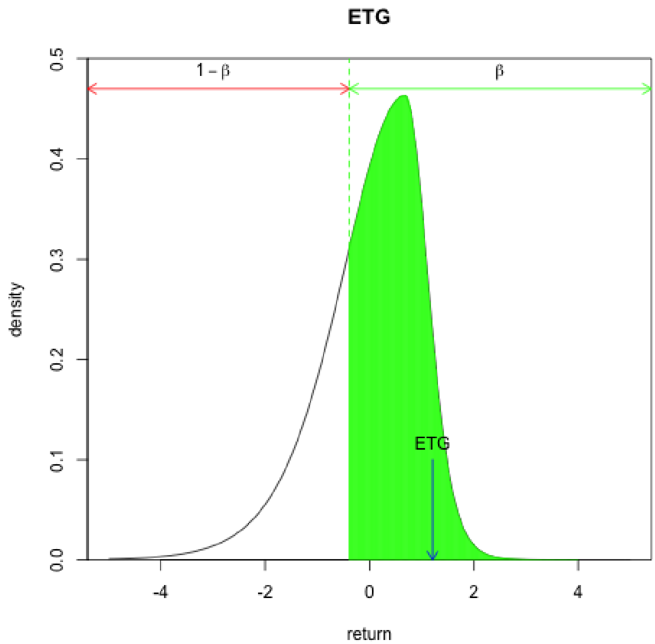

Figure 2.2: Illustration of ETG.

where β is the right tail probability. This thesis considers the reward measure corresponding to ETL, namely, the Expected Tail Gain (ETG).

Definition 2.2.1. The *Expected Tail Gain (ETG) at level* β, or *Conditional Expected Return (CER) at level* β, is defined as the expectation of the return conditioned on it being greater than the $(1-\beta)$-quantile:

$$\text{ETG}_\beta(r) = E[r|r \geq F^{-1}(1-\beta)].$$

As shown in Figure 2.2, β is the probability of the right tail. This corresponds to the definition of ETL except for a matter of sign. Specifically, we are interested in a relative large value for β (for example 90% or 95%) which includes the body of the distribution and a large right tail. By maximizing the CER associated with a large β, we are maximizing the average return in probablistic sense. When $\beta = 100\%$, CER shrinks to expected return.

2.3 Performance Measures (Ratios)

High returns are usually associated with high risk, and high risk investments are often associated with high returns. However, a one-time realized high return of a portfolio doesn't necessarily prove it to be a good strategy because it may be inherently risky. Several portfolio performance measures have been developed that attempt to balance the trade-off between returns and risk. Most of these measures take the form of a reward/risk ratio.

Sharpe Ratio

For a risky asset, the Sharpe ratio is defined as the excess return per unit of deviation/volatility,

$$\frac{\mu - r_f}{\sigma},$$

where μ is the expected return of the asset, r_f is the return on a benchmark asset,such as the risk free rate of return, σ is the volatility of the excess return. The Sharpe ratio characterizes how well the return of an asset compensates the investor for the risk taken. The higher the Sharpe ratio, the better.

Treynor Ratio

Treynor ratio is defined as the excess return per unit of market/systematic risk.

$$\frac{\mu - r_f}{\beta}.$$

Here β is the market risk coefficent in the capital asset pricing model.

The capital asset pricing model (CAPM) *If the market portfolio M is efficient, the expected return μ of any asset with random return r satisfies*

$$\mu - r_f = \beta(\mu_M - r_f)$$

where

$$\mu = E(r), \quad \mu_M = E(r_M), \quad \beta = \frac{\sigma_{rM}}{\sigma_M^2},$$

with $\sigma_{rM} = E((r - \mu_r)(r_M - \mu_M))$ is the covariance of r and r_M.

CAPM was developed by Treynor[45, 46], Sharpe[43], Lintner[22] and Mossin[28] independently. It relates the expected excess rate of return of an asset with the expected excess rate of return of the market portfolio through a proportionality factor β. The factor β can be interpreted as a normalized covariance of the asset with the market portfolio. If $\beta = 0$, that is, the asset is uncorrelated with the market, then $E(r) = r_f$. It seems contradictory that an asset with positive risk/volatility σ_r can have an expected return equal to the risk free rate. However, this is a consequence of the CAPM modeling assumption that the idiosyncratic risk of the asset can always be diversified away when combined with other assets to form a portfolio. Only the market risk represented by β contributes to the risk of the portfolio. This provides theoretical support for Treynor ratio which uses β as the risk measure in the denominator.

Jensen's Alpha (α)

Jensen's alpha is the excess return of an asset beyond its theoretical rate of return determined by CAPM:

$$\alpha = \mu - [r_f + \beta(\mu_M - r_f)].$$

Sortino Ratio

Sortino ratio [44, 13] is a modified version of Sharpe ratio. Instead of volatility , Sortino uses SSD as the risk measure in the denominator of the ratio,

$$\text{SOR} = \frac{\mu - r_f}{\text{SSD}}.$$

STARR Ratio

The STARR Ratio replaces volatility in the denominator of Sharpe ratio by VaR [26]:

$$STARR_\alpha = \frac{\mu - r_f}{\text{VaR}_\alpha(r)}.$$

Both Sortino ratio and STARR ratio are downside risk-adjusted performance measure. The next two ratios differ in the sense that they not only use downside risk measures in the denominator but also upside reward measures in the numerator.

Omega Ratio

Keating and Sadwick [7] proposed the Omega ratio as a performance measure. It is defined as

$$\Omega_c(r) = \frac{\int_c^\infty (1 - F(r))dr}{\int_{-\infty}^c F(r)dr},$$

where c is a threshold for the return. The above formula can also be written as

$$\Omega_c(r) = \frac{\int_c^\infty (r - c)f(r)dr}{\int_{-\infty}^c (c - r)f(r)dr} = \frac{P(r > c)E(r - c|r > c)}{P(r < c)E(c - r|r < c)}.$$

In this form Omega is seen as a probability weighted ratio of tail conditional expectations with fixed threshold c.

Rachev Ratio

Rachev ratio [26] uses expected tail loss (ETL) as the risk measure in the denominator and expected tail gain to measure the reward in the numerator:

$$RR_{\alpha,\beta}(r) = \frac{\text{ETL}_\beta(-r)}{\text{ETL}_\alpha(r)} = \frac{E(r|r \geq F^{-1}(1-\beta))}{-E(r|r \leq F^{-1}(\alpha))}.$$

Here α represents the left tail probability and β the right tail probability. By tuning the value of α and β, different portions of the distribution are used in the calculation of Rachev ratio. Possible choices are $\alpha = \beta = 5\%$ and $\alpha = 1\%, \beta = 50\%$. An empirical example can be found in Biglova et al.[1].

Note that the expected tail gain in the numerator is exactly the same as the definition of the Expected Tail Gain, or CER, given in Definition 2.2.1.

Chapter 3

PORTFOLIO OPTIMIZATION

In finance, a portfolio is a collection of financial assets such as bonds, stocks, funds and their derivatives. By investing in a portfolio, for a fixed return an investor is able to reduce the risk of the investment over any single component asset as long as they are not all perfectly correlated (correlation = 1). Portfolio optimization attempts to find the optimal weights of various assets that maximize the portfolio reward given a predetermined risk level, or alternatively, minimize the portfolio risk given a predetermined reward requirement. In this chapter, Section 3.1 briefly reviews those aspects of optimization theory of relevance to our discussion of portfolio optimization. In particular, the theory associated with the linear complementarity problem (LCP) and a primal-dual interior point method for its solution. Section 3.2 introduces the classical mean-variance portfolio optimization, also referred to as the modern portfolio theory (MPT). Section 3.3 discusses an adaptation of the MPT that uses ETL as the risk measure, develops an algorithm using primal-dual interior point method to solve this problem, and proves the equivalence of two alternative empirical formulation of this problem. Section 3.4 focuses on the ETG-ETL optimization problem. the core contribution of this dissertation. It employs ETG as the reward measure and ETL as the risk measure in a portfolio optimization framework.

3.1 Optimization Review

A standard constrained optimization framework takes the form

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, s \\ & f_i(x) = 0, \quad i = s+1, \ldots, m, \end{aligned} \tag{3.1.1}$$

where the vector $x = (x_1, \ldots, x_n)$ is the optimization variable, the function $f_0 : \mathbb{R}^n \to \mathbb{R}$ is called an *objective function*, and the functions $f_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \ldots, m$, are called *constraint functions*. A vector that satisfies all the constraints is called a *feasible solution*. An inequality constraint is said to be *active* at a feasible point x if $f_i(x) = 0$ and *inactive* if $f_i(x) < 0$ for some $i \in \{1, \ldots, s\}$. Equality constraints are always active at any feasible point. A feasible solution that minimizes the objective function is called an *optimal solution*, often denoted by x^*. The set of all optimal solutions are called the *optimal set*, and *optimal value* is defined as the greatest lower bound to the values of $f_0(x)$ as x ranges over the set of feasible solutions.

3.1.1 Some important classes of optimization problems

Optimization problems can be categorized into different classes based on the types of objective and constraint functions. Below we briefly describe three important classes for our study; linear programming, quadratic programming and convex programming.

Linear programming

The problem (3.1.1) is called a *linear program* (LP), if the objective and constraint functions $f_0, \ldots, f_m$ are all linear, i.e.

$$f_i(ax + by) = af_i(x) + bf_i(y) \tag{3.1.2}$$

for all $x, y \in \mathbb{R}^n$ and all $a, b \in \mathbb{R}$. A typical linear programming problem has the form

$$\begin{array}{ll} \text{minimize} & c^\mathsf{T} x \\ \text{subject to} & a_i^\mathsf{T} x \leq b_i, \quad i = 1, \ldots, m, \end{array} \tag{3.1.3}$$

where $c, a_1, \ldots, a_m \in \mathbb{R}^n$ and $b_1, \ldots, b_m \in \mathbb{R}$. Indeed, all linear programs can be transformed into one of this type. All the optimization problems that are not linear are called *nonlinear programming problems*.

Quadratic programming

The problem (3.1.1) is called a *quadratic program* (QP), if the objective function is quadratic and the constraint functions are linear. A quadratic program can be expressed in the form

$$\begin{array}{ll} \text{minimize} & \frac{1}{2} x^\mathsf{T} P x + q^\mathsf{T} x + r \\ \text{subject to} & Gx \leq h \\ & Ax = b, \end{array} \tag{3.1.4}$$

where $P \in \mathbb{S}_+^n, q \in \mathbb{R}^n, G \in \mathbb{R}^{m \times n}$, and $A \in \mathbb{R}^{p \times n}$.

Convex Optimization

A distinction between problems of the class (3.1.1) that is perhaps more important than linear and nonlinear, is *convex* and *non-convex*. A set $C \subset \mathbb{R}^n$ is said to be convex if

$$(1 - \lambda)x + \lambda y \in C \quad \forall\, x, y \in \mathbb{R}^n \text{ and } 0 \leq \lambda \leq 1 \ .$$

A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be convex if

$$f((1 - \lambda)x + \lambda y) \leq (1 - \lambda) f(x) + \lambda f(y) \quad \forall\, x, y \in \mathbb{R}^n \text{ and } 0 \leq \lambda \leq 1 \ .$$

A function $g : \mathbb{R}^n \to \mathbb{R}$ is said to be concave if $-g$ is convex. A typical convex optimization problem has the form

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le b_i, \quad i = 1, \ldots, m, \\ & Ax = b, \end{aligned} \tag{3.1.5}$$

where functions $f_0, \ldots, f_m$ are all convex. Clearly, linear functionals are convex, and so LP's are examples of convex programs. A QP is convex if and only if the matrix P in (3.1.4) is positive semi-definite. All the optimization problems that are not convex are called *non-convex programming problems.* The ETG-ETL optimization problem considered in Section 3.4 is a non-convex programming problem. Specifically, it is called a *concave programming problem* which is the class of problems where the objective function is concave while the constraint region is convex.

3.1.2 The Lagrangian and duality

Consider the standard optimization problem

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, s \\ & h_i(x) = 0, \quad i = 1, \ldots, m, \end{aligned} \tag{3.1.6}$$

where $x \in \mathbb{R}^n$. Denote the feasible set by

$$\mathcal{D} := \{x \in \mathbb{R}^n \ : \ f_i(x) \le 0, \ i = 1, \ldots, s\text{and } h_i(x) = 0, \ i = 1, \ldots, m\} \ .$$

Definition 3.1.1. The *Lagrangian* $L : \mathbb{R}^n \times \mathbb{R}^s \times \mathbb{R}^m \to \mathbb{R}$ for the optimization problem (3.1.6) is given by

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{s} \lambda_i f_i(x) + \sum_{i=1}^{m} \nu_i h_i(x) \tag{3.1.7}$$

with $\mathrm{dom}(L) = \mathcal{D} \times \mathbb{R}^s_+ \times \mathbb{R}^m$, where λ_i's and ν_i's are called the *Lagrange multipliers* associated with corresponding constraints.

Definition 3.1.2. The *dual objective function* $g : \mathbb{R}^s \times \mathbb{R}^m \to \mathbb{R}$ is the minimum value of the Lagrangian over x: for $\lambda \in \mathbb{R}^s$, $\nu \in \mathbb{R}^m$,

$$\begin{aligned} g(\lambda, \nu) &:= \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \\ &= \inf_{x \in \mathcal{D}} f_0(x) + \sum_{i=1}^{s} \lambda_i f_i(x) + \sum_{i=1}^{m} \nu_i h_i(x) \ . \end{aligned}$$

If we denote the optimal value of problem (3.1.6) by p^*, it is easy to check that $g(\lambda, \nu) \leq p^*$ for any $\lambda \geq 0$ and $\nu \in \mathbb{R}^m$, i.e., $g(\lambda, \nu)$ provides a lower bound for p^*.

Definition 3.1.3. The *Lagrangian dual problem* to the problem (3.1.6) is the optimization problem

$$\begin{aligned} &\text{maximize} \quad g(\lambda, \nu) \\ &\text{subject to} \quad \lambda \geq 0 \end{aligned} \tag{3.1.8}$$

In this context, problem (3.1.6) is often referred to as the *primal problem* and (3.1.8) as the *dual problem.* The dual problem is a convex programming problem regardless of whether the primal problem is convex or not. If we let d^* denote the optimal value for (3.1.8), we have

$$d^* \leq p^* \ .$$

This is called *weak duality.* If in fact

$$d^* = p^* \ ,$$

we say *strong duality* holds. Strong duality does not hold in general, and does not necessarily imply that solutions to either the primal or dual problem exist. In convex programming, strong duality can usually be establish by showing that a *constraint qualification* is satisfied.

Consider a convex optimization problem in the form of (3.1.5), a simple constraint qualification is the *Slater's condition*: there exists an $x \in \mathcal{D}$ such that

$$f_i(x) < b_i, \quad , i = 1, \ldots, m, \qquad Ax = b.$$

In other word, the inequality constraints in (3.1.5) are strictly satisfied. *Slater's Theorem* states that for a convex problem Slater's condition guarantees strong duality as well as the existence of a solution to the dual problem. For linear programming problems, strong duality always holds and solution to both the primal and dual problems always exist as long as the primal and dual problems are both feasible.

3.1.3 KKT optimality conditions

Again consider the standard optimization problem (3.1.6). We assume $f_0, \ldots, f_s$, $h_1, \ldots, h_m$ are differentiable. Suppose strong duality holds, x^* is the optimal solution to the primal problem and (λ^*, ν^*) is the optimal solution to the dual problem. Then the following conditions hold:

$$\begin{aligned} f_i(x^*) &\leq 0, \quad i = 1, \ldots, s \\ h_i(x^*) &= 0, \quad i = 1, \ldots, m \\ \lambda_i^* &\geq 0, \quad i = 1, \ldots, s \\ \lambda_i^* f_i(x^*) &= 0, \quad i = 1, \ldots, s \\ \nabla f_0(x^*) + \sum_{i=1}^{s} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{m} \nu_i^* \nabla h_i(x^*) &= 0. \end{aligned}$$

These are called the *Karush-Kuhn-Tucker* (KKT) conditions. If the primal problem is convex, then any point $(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$ satisfying the KKT conditions is such that $\tilde{x}$ is an optimal solution to the primal problem and (λ^*, ν^*) is an optimal solution to the dual problem. For this reason, the KKT conditions are often used to find the optimal solutions to optimization problems.

3.1.4 The linear complementarity problem

Given a real matrix $M \in \mathbb{R}^{n \times n}$ and vector $q \in \mathbb{R}^n$, the *linear complementarity problem* (LCP) is: Find $x, y \in \mathbb{R}^n$ such that

$$y = Mx + q, \quad x^\mathsf{T} y = 0, \quad 0 \leq x, \quad 0 \leq y \ . \tag{3.1.9}$$

The KKT conditions for the QP

$$\begin{aligned} \text{minimize} \quad & f(x) = c^{\mathsf{T}}u + \tfrac{1}{2}u^{\mathsf{T}}Qu \\ \text{subject to} \quad & Au \leq b \\ & u \geq 0, \end{aligned} \tag{3.1.10}$$

can stated in the form of an LCP by setting

$$M = \begin{bmatrix} Q & A^{\mathsf{T}} \\ -A & 0 \end{bmatrix} \quad \text{and} \quad q = \begin{bmatrix} c \\ b \end{bmatrix}.$$

3.1.5 Primal-dual interior point method

In this section we provide a sketch of the basic ideas behind the primal-dual interior point method for solving the LCP problem. Define

$$F(x, y) = \begin{bmatrix} Mx - y + q \\ XYe \end{bmatrix}, \tag{3.1.11}$$

where X and Y are the diagonal matrices whose diagonal elements are those of the vectors x and y, respectively. Then solving the LCP problem (3.1.9) is equivalent to solving the equation

$$F(x, y) = 0, \quad \text{for some } x \geq 0 \text{ and } y \geq 0. \tag{3.1.12}$$

Therefore, we can solve the LCP problem by applying variants of Newton's method to the function (3.1.11) modifying the search directions and step lengths so that the inequalities $x \geq 0$ and $y \geq 0$ are satisfied *strictly* at every iteration. First, we derive the Newton step and then show how to correct it by introducing the concept of the *central path*. From step k to $k+1$, the first order approximation to $F(z^{k+1})$ at $z = (x^{\mathsf{T}}, y^{\mathsf{T}})^{\mathsf{T}}$ isgiven by

$$F(z^{k+1}) = F(z^k) + J_F(z^k)(z^{k+1} - z^k),$$

where $J_F(\cdot)$ is the Jacobian matrix of F. Setting the left side zero and solve for Newton's direction d_N^k, we get

$$d_N^k = z^{k+1} - z^k = -J_f(z^k)^{-1}F(z^k).$$

If we take a full step along Newton's direction at each iteration, the values for (x, y) could easily go beyond the boundary of positive quadrant and violate the constraints $x, y \geq 0$ before reaching the solution. Instead we take a shorter step $z^{k+1} = z^k + \alpha d_N^k$, where $\alpha \in (0, 1]$. However, if α is too small, progress toward a solution will be insufficient to assure convergence. To address this problem, we solve the modified subproblem

$$F(x, y) = \begin{bmatrix} Mx - y + q \\ XYe \end{bmatrix} = \begin{bmatrix} 0 \\ te \end{bmatrix}, \quad x \geq 0, \quad y \geq 0, \tag{3.1.13}$$

for $t > 0$ where $e \in \mathbb{R}^n$ is the vector of all ones. The set of all solutions to such problems for $t > 0$ is called the *central path* for the LCP problem:

$$\mathcal{C} := \{z \, : \, F(z) = (0 \; e^{\mathsf{T}})^{\mathsf{T}}, \; z > 0\} \, .$$

A modified Newton step is then computed by solving for points on the central path:

$$\begin{pmatrix} \mathbf{0} \\ \sigma\tau e \end{pmatrix} = F(z^{k+1}) = F(z^k) + J_F(z^k)(z^{k+1} - z^k), \tag{3.1.14}$$

$$d^k = z^{k+1} - z^k = -J_F(z^k)^{-1}\left[F(z^k) - \begin{pmatrix} \mathbf{0} \\ \sigma\tau e \end{pmatrix}\right]. \tag{3.1.15}$$

In this equation, $\tau = (x^{\mathsf{T}}y)/n$ is defined as the *duality measure*, which is the average value of the pairwise products $x_i y_i$, $i = 1, \ldots, n$. Instead of directly aiming at $\tau = 0$ as in the previous case, we first want to achieve a fraction σ of the duality measure at this step, where $\sigma \in [0, 1]$ is called the *centering parameter*. If $\sigma = 0$, d^k is just the Newton's direction; if $\sigma > 0$, it is possible to take a longer stepsize α along the direction before violating the nonnegative constraints because we are aiming towards more towards the central path point (x, y) with $XYe = \tau e$. Therefore, by shrinking values for σ carefully and choosing a suitable step-size α at each iteration, we are able to achieve the global convergence and polynomial complexity of the method.

3.2 Mean-Variance Portfolio Optimization

Harry Markowitz[24, 25] introduced modern portfolio theory in the 1950s. He used volitility as a proxy for the portfolio risk and expected return as a proxy of the portfolio reward. There are two equivalent forms of the mean-variance portfolio optimization problem:

1. maximize the portfolio's expected return for a given level of risk;

2. minimize the portfolio risk for a given level of expected return.

This section develops the mathematical model and shows that the second form can be represented by a quadratic programming problem.

For a portfolio consisting of n assets, let r_i be the return of asset i and let w_i be the amount of capital invested in asset i for $i = 1, 2, \ldots, n$. Without loss of generality, we take the total principal to be \$1, so that

$$w^{\mathsf{T}} e = 1, \tag{3.2.1}$$

where $w = (w_1, \ldots, w_n)^{\mathsf{T}}$, $e = (1, \ldots, 1)^{\mathsf{T}}$. Then the expected return of the portfolio, r_p, is given by the weighted average of the return of each asset,

$$r_p = \sum_{i=1}^{n} w_i r_i = w^{\mathsf{T}} r. \tag{3.2.2}$$

Here $r = (r_1, \ldots, r_n)^{\mathsf{T}}$. Each r_i is a random variable, so for each fixed w, r_p, as a linear combination of r_i's, is also a random variable. Let $\mu_i = E(r_i)$ denote the expected return of asset i, then the expected return of the portfolio is the weighted average of the expected return for each asset,

$$\mu_p = E(r_p) = E\left(w^{\mathsf{T}} r\right) = w^{\mathsf{T}} \mu, \tag{3.2.3}$$

where $\mu = (\mu_1, \dots, \mu_n)^\mathsf{T}$.

Assume the investor has a target expected return of ρ This target can be expressed as the linear constraint

$$w^\mathsf{T}\mu \geq \rho. \tag{3.2.4}$$

The variance of the portfolio return is given by

$$\begin{aligned}
\sigma_p^2 &= E(r_p - \mu_p)^2 &(3.2.5)\\
&= E(w^\mathsf{T} r - w^\mathsf{T}\mu)^2 &(3.2.6)\\
&= E[w^\mathsf{T}(r-\mu)(r-\mu)^\mathsf{T} w] &(3.2.7)\\
&= w^\mathsf{T}\Omega w. &(3.2.8)
\end{aligned}$$

where $\Omega = E[(r-\mu)(r-\mu)^\mathsf{T}]$ is the covariance matrix of the n assets. The classical Markowitz mean-variance portfolio optimization can be written as

$$\begin{aligned}
\underset{w\in\mathbb{R}^n}{\text{minimize}} \quad & w^\mathsf{T}\Omega w &(3.2.9)\\
\text{subject to} \quad & w^\mathsf{T}\mu = \rho &(3.2.10)\\
& w^\mathsf{T} e = 1 &(3.2.11)
\end{aligned}$$

Notice that variance, not volitility, is used as the objective function to minimize. This makes the problem a convex quadratic programming problem which is easy to solve and gives the same optimal solution w^* as that given by a volitility objective. Also note that the inequality constraint in (3.2.4) is replaced by an equality constraint in (3.2.10). This does not change the optimal solution because it is an active constraint at the solution, i.e. equality must hold at any optimal solution w^*. Using constraint (3.2.10) enables us to transform the problem into a more favorable format, which is explored in detail later.

In general, we assume that Ω is positive definite. This is equivalent to assuming that there there does not exist a non-trivial linear combination of the assets $r_p = w_1 r_1 + \cdots + w_n r_n$ such that $P(r_p = \text{constant}) = 1$. Intuitively, there should not exist a linear combination of the assets that yields a deterministic return. If this assumption is met, the solution to the Markowitz mean-variance portfolio optimization problem (3.2.9) is unique when it exists.

To make use of this model in portfolio selection, we must specify both the vector expected returns μ and the covariance matrix Ω. There are several options for the estimation of μ and Ω. A straightforward approach is to use the empirical non-parametric estimation, i.e. take the mean of historical returns as the expected return, and take the variance of the historical data as an estimate of Ω. One shortfall of this method is that it requires $n(n+1)/2$ estimators for the covariance matrix and n for μ. Another issue is the stability of the estimation since it can differ greatly between similar historical periods and is easily affected by outliers. Another approach is to assume each r_i follows a known distribution function with explicit density function, such as the normal distribution or a skewed t distritution. One then first fits the paramterized distribution to the historical data using maximum likelihood estimation to estimate the values of the parameters, then calculate expected returns and covariance using the explicit density function. While using this approach, asymmetric distributions such as skewed t distributions are preferred over the normal distribution since they capture the fat tail and skewness of the returns detected in empirical research. A third option is to make use of factor models assuming the returns are driven by some factor returns. The factor candidates can be macroeconomic factors such as index returns and inflation rate, fundamental factors such as P/E ratio and returns on a model portfolio, and statistical factors such as principal components. Factor models can reduce the number of estimations and give more robust results. If there are p factors, then we only need $np + n + p^2$ estimations for the covariance matrix Ω.

3.3 Mean-ETL Portfolio Optimization

3.3.1 Formulation of the mean-ETL optimization problem

As discussed in Section 2.1.2, the expected tail loss ETL, or CVAR, is a coherent risk measure and so is preferable to volatility. A natural extension of classical Markowitz portfolio optimization problem is to replace variance by ETL as the objective function to minimize:

$$\begin{aligned} \text{minimize}_{w\in\mathbb{R}^n} \quad & \text{ETL}_\alpha(w) = -E[w^\mathsf{T} r | w^\mathsf{T} r \leq -\text{VaR}_\alpha(w^\mathsf{T} r)] \\ \text{subject to} \quad & w^\mathsf{T}\mu = \rho \\ & w^\mathsf{T} e = 1 \ . \end{aligned} \tag{3.3.1}$$

In this form, this problem is difficult to solve particularly since the definition of ETL involves VaR. Rockafellar and Uryasev [37, 38] developed an equivalent formulation of ETL by expressing it as the optimal value of a minimization problem. Let $r_p = w^\mathsf{T} r$ denote the portfolio return function, where w is a deterministic vector variable representing the portfolio weights and r is a vector random variable following some distribution. For each fixed w, r_p is a scalar random variable whose distribution can be derived from that of r. The cumulative distribution function associated with w is

$$\Phi(w, z) = \int_{r_p \leq z} p(r)dr \tag{3.3.2}$$

where $p(r)$ is the density function of r. For $\alpha \in (0, 1)$, the α-VaR and α-ETL associated with w is

$$\text{VaR}_\alpha(w) = -\inf\{z | \Phi(w, z) \geq \alpha\} \qquad \text{and} \tag{3.3.3}$$

$$\text{ETL}_\alpha(w) = -\frac{1}{\alpha} \int_{r_p \geq \text{VaR}_\alpha(w)} r_p p(r)dr \ , \tag{3.3.4}$$

respectively. Now define a new function

$$F_\alpha(w, C) = C + \frac{1}{\alpha} \int [-r_p - C]_+ p(r)dr \ , \tag{3.3.5}$$

where $z_+ = max\{0, z\}$, $z_- = max\{0, -z\}$, and C is a scalar variable.

Theorem 3.3.1. *[37, Theorem 1] $F_\alpha(w, C)$ as a function of C is convex and continuously differentiable. The α-ETL associated with w can be obtained by solving*

$$ETL_\alpha(w) = \min_{C \in R} F_\alpha(w, C) \tag{3.3.6}$$

Also the set of values of C, where the minimum is attained, i.e.

$$A_\alpha(w) = \arg\min_{C \in \mathbb{R}} F_\alpha(w, C), \tag{3.3.7}$$

is a nonempty, closed, and bounded interval satisfying

$$VaR_\alpha(w) = \textit{left endpoint of } A_\alpha(w). \tag{3.3.8}$$

Observe that minimizing α-ETL over $w \in W$ is equivalent to minimizing $F_\alpha(w, C)$ over all $(w, C) \in W \times \mathbb{R}$, i.e.

$$\min_{w \in W} \mathrm{ETL}_\alpha(w) = \min_{(w,r) \in W \times \mathbb{R}} F(w, C) \tag{3.3.9}$$

where $W \subset \mathbb{R}^n$ is the feasible set for w, which in our mean-ETL optimization problem is a *polyhedral* convex set. Typically, $A_\alpha(w)$ shrinks to a single point, so by directly solving the minimization problem on the right side in (3.3.9), we get the optimal weights giving the minimum ETL, denoted by w^*, and the corresponding VaR value denoted by C^*, i.e.

$$\min_{w \in W} \mathrm{ETL}_\alpha(w) = F(w^*, C^*) \qquad \mathrm{VaR}_\alpha(w^*) = C^* \ .$$

According to (3.3.9), the mean-ETL optimization problem can be formulated as

$$\begin{aligned} \underset{w,C}{\text{minimize}} \quad & F(w, C) = C + \frac{1}{\alpha} \int [-r_p - C]_+ p(r) dr \\ \text{subject to} \quad & w^\mathsf{T} \mu = \rho \quad w^\mathsf{T} e = 1 \ . \end{aligned} \tag{3.3.10}$$

Assume there is a historical realization of r, denoted by $r^1, r^2, \ldots, r^T$ in T periods, and assume equal probability of each historical return, then $F(w, C)$ can be approximated by

$$\hat{F}(w, C) = C + \frac{1}{\alpha T} \sum_{t=1}^{T} [-w^\mathsf{T} r^t - C]_+ \tag{3.3.11}$$

Then this empirical estimation $\hat{F}(w, C)$ is used as the target function in the mean-ETL portfolio optimization problem (3.3.10). Since $\hat{F}(w, C)$ is piecewise linear and convex in w and C and all the constraints are linear, this is a convex programming problem. By introducing a vector variable $z = (z_1, \ldots, z_T)^\mathsf{T}$, we can rewrite it as a linear programming problem,

$$\begin{aligned}
\underset{w,C,z}{\text{minimize}} \quad & C + \beta z^\mathsf{T} e && && (3.3.12)\\
\text{subject to} \quad & 0 \le z + Rw + Ce && u && (3.3.13)\\
& 0 \le z && v && (3.3.14)\\
& w^\mathsf{T}\mu = \rho && \gamma && (3.3.15)\\
& w^\mathsf{T} e = 1 && \lambda && (3.3.16)
\end{aligned}$$

where $\beta = 1/(\alpha T)$, $R = (r^1, \ldots, r^T)^\mathsf{T} \in \mathbb{R}^{T\times n}$. As in the case of Markowitz mean-variance portfolio optimization problem (3.2.9), we assume that there is no non-trivial linear combination of the assets $r_p = w_1 r_1 + \cdots + w_n r_n$ such that $P(r_p = \text{constant}) = 1$. We instantiate this condition in our finite sample of these assets by making the following assumption:

$$\nexists\, w \in \mathbb{R}^n,\ \kappa \in \mathbb{R} \quad \text{such that } Rw = \kappa e\ . \tag{3.3.17}$$

In the next section, we implement the primal-dual interior point method to solve (3.3.10).

3.3.2 Primal-Dual interior point algorithm

This section starts with finding the dual problem and complimentarity conditions of problem (3.3.12)-(3.3.16). Let $u \in \mathbb{R}^T$, $v \in \mathbb{R}^T$, $\gamma \in \mathbb{R}$ and $\lambda \in \mathbb{R}$ be the corresponding lagrangian multipliers for constraints (3.3.13)-(3.3.16), then the lagarangian is

$$\begin{aligned} L(C, z, w; u, v, \gamma, \lambda) &= C + \beta z^{\mathsf{T}} e - u^{\mathsf{T}}(z + Rw + Ce) - v^{\mathsf{T}} z \\ &\quad + \gamma(\mu^{\mathsf{T}} w - \rho) + \lambda(w^{\mathsf{T}} e - 1) && (3.3.18) \\ &= C(1 - u^{\mathsf{T}} e) + z^{\mathsf{T}}(\beta e - u - v) \\ &\quad + w^{\mathsf{T}}(-R^{\mathsf{T}} u + \gamma\mu + \lambda e) - \rho\gamma - \lambda && (3.3.19) \end{aligned}$$

Take the partial derivative of L with respect to C, z and w respectively, and set each to 0, we get

$$\frac{\partial L}{\partial C} = 1 - u^{\mathsf{T}} e = 0 \tag{3.3.20}$$

$$\nabla_z L = \beta e - u - v = 0 \tag{3.3.21}$$

$$\nabla_w L = -R^{\mathsf{T}} u + \gamma\mu + \lambda e = 0 \tag{3.3.22}$$

which leads to the constraints of the dual problem. Gather everything together, the dual problem is

$$\begin{aligned} \underset{u,v,\gamma,\lambda}{\text{minimize}} \quad & \rho\gamma + \lambda && (3.3.23) \\ \text{subject to} \quad & u \geq 0, \quad v \geq 0 && (3.3.24) \\ & u^{\mathsf{T}} e = 1 && (3.3.25) \\ & \beta e - u - v = 0 && (3.3.26) \\ & R^{\mathsf{T}} u - \gamma\mu - \lambda e = 0 && (3.3.27) \end{aligned}$$

Also the KKT conditions are

$$\text{Primal feasibility} \quad \begin{cases} 0 \le z \\ 0 \le z + Rw + Ce \\ \nu^\mathsf{T} w = \rho \\ e^\mathsf{T} w = 1 \end{cases} \tag{3.3.28}$$

$$\text{Dual feasibility} \quad \begin{cases} u \ge 0, \quad v \ge 0 \\ u^\mathsf{T} e = 1 \\ \beta e - u - v = 0 \\ R^\mathsf{T} u - \gamma\mu - \lambda e = 0 \end{cases} \tag{3.3.29}$$

$$\text{Complementary slackness} \quad \begin{cases} u^\mathsf{T}(z + Rw + Ce) = 0 \\ v^\mathsf{T} z = 0 \end{cases} \tag{3.3.30}$$

Strong duality for linear programming always holds as long as both the primal and dual problems are feasible. In addition, for convex optimization problems, if the Slater constraint qualification holds, the KKT conditions provide necessary and sufficient conditions for optimality. Therefore, in order to solve (3.3.12)(3.3.16) it suffices to solve the KKT conditions to obtain optimal solutions to both the primal and dual problems when they exist. Define the function $F : \mathbb{R}^{3+4T+N} \to \mathbb{R}^{3+4T+N}$ as follows:

$$F(z, y, u, v, C, w, \lambda, \gamma) = \begin{bmatrix} Rw + Ce + z - y \\ w^\mathsf{T}\mu - \rho \\ w^\mathsf{T} e - 1 \\ R^\mathsf{T} u - \gamma\mu - \lambda e \\ u + v - \beta e \\ e^\mathsf{T} u - 1 \\ YUe \\ VZe \end{bmatrix},$$

where Y is the diagonal matrix having the elements of vector y on the diagonal, with the diagonal matrices U, V and Z defined similarly. Clearly, solving the KKT conditions is equivalent to solving the system:

$$F(z, y, u, v, C, w, \lambda, \gamma) = \mathbf{0} \tag{3.3.31}$$

$$(z, y, u, v) \geq \mathbf{0}. \tag{3.3.32}$$

We can use the previously introduced primal-dual interior point method for this purpose. The following is a derivation of the Jacobian matrix of F:

$$J_F = \begin{array}{c} \\ VZe \\ YUe \\ u+v-\beta e \\ Rw + Ce + z - y \\ e^{\mathsf{T}}u - 1 \\ R^{\mathsf{T}}u - \gamma\mu - \lambda e \\ w^{\mathsf{T}}e - 1 \\ w^{\mathsf{T}}\mu - \rho \end{array} \begin{array}{c} \begin{array}{cccccccc} z & y & u & v & C & w & \lambda & \gamma \end{array} \\ \begin{bmatrix} V & 0 & 0 & Z & 0 & 0 & 0 & 0 \\ 0 & U & Y & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & I & 0 & 0 & 0 & 0 \\ I & -I & 0 & 0 & e & R & 0 & 0 \\ 0 & 0 & e^{\mathsf{T}} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & R^{\mathsf{T}} & 0 & 0 & 0 & -e & -\mu \\ 0 & 0 & 0 & 0 & 0 & e^{\mathsf{T}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu^{\mathsf{T}} & 0 & 0 \end{bmatrix} \end{array}$$

We find its block LU decomposition by Gaussian elimination under the assumption (3.3.17) on the matrix R. This is used to calculate its inverse in finding the modified Newton's direction later. The Gaussian Elimination proceeds as follows:

$$\Rightarrow \begin{array}{c} \text{row1} \\ \\ \\ -V^{-1}\text{row1} \\ \\ \end{array} \begin{bmatrix} V & 0 & 0 & Z & 0 & 0 & 0 & 0 \\ 0 & U & Y & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & I & 0 & 0 & 0 & 0 \\ 0 & -I & 0 & -V^{-1}Z & e & R & 0 & 0 \\ & & & \cdots & & & & \end{bmatrix}$$

$$
\begin{array}{r}
\\
\text{row2} \\
\\
\Rightarrow +U^{-1}\text{row2} \\
\\
\\
\\
\end{array}
\begin{bmatrix}
V & 0 & 0 & Z & 0 & 0 & 0 & 0 \\
0 & U & Y & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & I & I & 0 & 0 & 0 & 0 \\
0 & 0 & U^{-1}Y & -V^{-1}Z & e & R & 0 & 0 \\
0 & 0 & e^{\mathsf{T}} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & R^{\mathsf{T}} & 0 & 0 & 0 & -e & -\mu \\
 & & & \cdots & & & &
\end{bmatrix}
$$

$$
\begin{array}{c}
\\
\\
\text{row3} \\
\Rightarrow -U^{-1}Y\text{row3}, \times(-1) \\
-e^{\mathsf{T}}\text{row3} \\
-R^{\mathsf{T}}\text{row3} \\
\\
\end{array}
\begin{bmatrix}
V & 0 & 0 & Z & 0 & 0 & 0 & 0 \\
0 & U & Y & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & I & I & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & V^{-1}Z + U^{-1}Y & -e & -R & 0 & 0 \\
0 & 0 & 0 & -e^{\mathsf{T}} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -R^{\mathsf{T}} & 0 & 0 & -e & -\mu \\
 & & & \cdots & & & &
\end{bmatrix}
$$

Let $D = V^{-1}Z + U^{-1}Y$,

$$
\begin{array}{c}
\\
\\
\\
\Rightarrow \qquad \text{row4} \\
+e^{\mathsf{T}}D^{-1}\text{row4}, \times(-1) \\
+R^{\mathsf{T}}D^{-1}\text{row4}, \times(-1) \\
\\
\end{array}
\begin{bmatrix}
V & 0 & 0 & Z & 0 & 0 & 0 & 0 \\
0 & U & Y & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & I & I & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & D & -e & -R & 0 & 0 \\
0 & 0 & 0 & 0 & e^{\mathsf{T}}D^{-1}e & e^{\mathsf{T}}D^{-1}R & 0 & 0 \\
0 & 0 & 0 & 0 & R^{\mathsf{T}}D^{-1}e & R^{\mathsf{T}}D^{-1}R & e & \mu \\
 & & & \cdots & & & &
\end{bmatrix}
$$

Let $\eta = e^\top D^{-1} e$,

$$\Rightarrow \quad \begin{array}{r} \\ \\ \\ \\ \text{row5} \\ -\frac{1}{\eta}R^\top D^{-1}e\text{row5} \\ \\ \\ \end{array} \begin{bmatrix} V & 0 & 0 & Z & 0 & 0 & 0 & 0 \\ 0 & U & Y & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & D & -e & -R & 0 & 0 \\ 0 & 0 & 0 & 0 & \eta & e^\top D^{-1}R & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \Gamma & e & \mu \\ 0 & 0 & 0 & 0 & 0 & e^\top & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu^\top & 0 & 0 \end{bmatrix}$$

Let $\Gamma = R^\top D^{-1}R - \frac{1}{\eta}R^\top D^{-1}ee^\top D^{-1}R$. Note that Γ is invertible. Since it can be written as

$$\Gamma = R^\top D^{-1}(D - \frac{1}{\eta}ee^\top)D^{-1}R. \tag{3.3.33}$$

In the above equation, the rank of $(D - \frac{1}{\eta}ee^\top)D^{-1}$ is $T-1$ and e is the base of its null space because

$$(D - \frac{1}{\eta}ee^\top)D^{-1}e = e - e = 0 \tag{3.3.34}$$

Under the assumption (3.3.17), e is not in the range of R, so for any nonzero $w \in \mathbb{R}^n$,

$$w^\top \Gamma w \neq 0. \tag{3.3.35}$$

Hence, Γ is invertible. Then in the next step, we have

$$\Rightarrow \begin{array}{r} \\ \\ \\ \\ \\ \text{row6} \\ -e^{\top}\Gamma^{-1}\text{row6}, \times(-1) \\ -\mu^{\top}\Gamma^{-1}\text{row6}, \times(-1) \end{array} \begin{bmatrix} V & 0 & 0 & Z & 0 & 0 & 0 & 0 \\ 0 & U & Y & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & D & -e & -R & 0 & 0 \\ 0 & 0 & 0 & 0 & \eta & e^{\top}D^{-1}R & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \Gamma & e & \mu \\ 0 & 0 & 0 & 0 & 0 & 0 & e^{\top}\Gamma^{-1}e & e^{\top}\Gamma^{-1}\mu \\ 0 & 0 & 0 & 0 & 0 & 0 & \mu^{\top}\Gamma^{-1}e & \mu^{\top}\Gamma^{-1}\mu \end{bmatrix}$$

$$\Rightarrow L = \begin{bmatrix} V & 0 & 0 & Z & 0 & 0 & 0 & 0 \\ 0 & U & Y & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & D & -e & -R & 0 & 0 \\ 0 & 0 & 0 & 0 & \eta & e^{\top}D^{-1}R & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \Gamma & e & \mu \\ 0 & 0 & 0 & 0 & 0 & 0 & e^{\top}\Gamma^{-1}e & e^{\top}\Gamma^{-1}\mu \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \delta \end{bmatrix}$$

where $\delta = \mu^{\top}\Gamma^{-1}\mu - \frac{1}{e^{\top}\Gamma^{-1}e}\mu^{\top}\Gamma^{-1}ee^{\top}\Gamma^{-1}\mu$. Simultaneously, we can construct the lower-triangle matrix by doing the following operations to an identiry matrix as follows

$$\begin{bmatrix} I_{T\times T} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I_{T\times T} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I_{T\times T} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{T\times T} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I_{n\times n} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Operations:

row8 $+\frac{1}{e^\mathsf{T}\Gamma^{-1}e}\mu^\mathsf{T}\Gamma^{-1}e$ row7, $\times(-1)$ then $+\mu^\mathsf{T}\Gamma^{-1}$row6;

row7 $\times(-1)$, $+e^\mathsf{T}\Gamma^{-1}$row6;

row6 $+\frac{1}{\eta}R^\mathsf{T}D^{-1}e$ row5, $\times(-1)$, $-R^\mathsf{T}D^{-1}$row4, $+R^\mathsf{T}$row3;

row5 $\times(-1)$,$-e^\mathsf{T}D^{-1}$row4, $+e^\mathsf{T}$row3;

row4 $\times(-1)$, $+U^{-1}Y$row3, $-U^{-1}$row2, $+V^{-1}$row1

Denote the resulting block lower-triangle matrix by P, then we have a block LU decomposition for J_F: $J_F = PL$. This makes it efficient to compute the inverse of the Jacobian of F in (3.1.15) to solve for the modified Newton's direction d^k.

3.3.3 An alternative formulation

In Section 3.3.1, we used the empirical approximation to the mean-ETL optimization problem (3.3.10). An alternative approach is to use the empirical approximation to problem (3.3.1), which uses the original definition of ETL. This section formulates the second problem and compares the difference between the two approximations. For easy reference, we restate the two problems below. Given a historical realization of vector returns r: $r^1, r^2, \ldots, r^T$ in T periods, which we assume occur with equal probability, the approximate mean-ETL optimization problems are

$$\begin{aligned} \underset{w,C}{\text{minimize}} \quad & \hat{F}(w,C) = C + \frac{1}{\alpha T}\sum_{t=1}^{T}[-w^\mathsf{T}r^t - C]_+ \\ \text{subject to} \quad & w^\mathsf{T}\mu = \rho \quad w^\mathsf{T}e = 1, \end{aligned} \tag{$\mathcal{P}_1$}$$

and

$$\begin{aligned} \underset{w}{\text{minimize}} \quad & \widehat{\mathrm{ETL}}_\alpha(w) = -\frac{1}{\lfloor \alpha T \rfloor} \sum_{t=1}^{\lfloor \alpha T \rfloor} (Rw)_{(t)} \\ \text{subject to} \quad & w^\mathsf{T} \mu = \rho \quad w^\mathsf{T} e = 1. \end{aligned} \tag{$\mathcal{P}_2$}$$

where $R = (r^1, \ldots, r^T)^\mathsf{T}$ and $(Rw)_{(t)}$ is the tth smallest element of Rw. With value of$w$ fixed, let $z_t = w^\mathsf{T} r^t$ for $t = 1, \ldots, T$. Then z_t is the portfolio's return in period t. In this section we show that the two approximating problems $\mathcal{P}_1$ and $\mathcal{P}_2$ are the same when αT is an integer. The proof begins with three lemmas. Define

$$G(C) = C + \frac{1}{N} \sum_{t=1}^{T} [-z_t - C]_+ \tag{3.3.36}$$

where $N \in [1, T]$.

Lemma 3.3.2. *If $N \in [1, T]$ is an integer, then*

$$\arg\min_C G(C) = [-z_{(N+1)}, -z_{(N)}] \ ,$$

where $z_{(k)}$ is the kth smallest element of the vector z.

Proof. It is easy to check that $G(C)$ is a piecewise linear convex function on $C \in \mathbb{R}$. To prove this lemma, it suffices to show that 0 is in its subdifferential only when $C \in [-z_{(N+1)}, -z_{(N)}]$. The subdifferential of $G(C)$ is

$$\partial G(C) = 1 + \frac{1}{N} \sum_{C < -z_t} (-1) + \frac{1}{N} \sum_{C > -z_t} 0 + \frac{1}{N} \sum_{C = -z_t} [-1, 0] \ .$$

Case 1. If $C < -z_{(N+1)}$, then $C < -z_{(k)}$ for $k = 1, \ldots, N+1$ so

$$1 - \frac{T}{N} \leq \partial G(C) \leq 1 - \frac{N+1}{N} < 0 \quad \Rightarrow \quad 0 \notin \partial G(C) \ .$$

Case 2. If $C > -z_{(N)}$, then $C > -z_{(k)}$ for $k = N, \ldots, T$ so

$$0 < 1 - \frac{N-1}{N} \leq \partial G(C) < 1 \quad \Rightarrow \quad 0 \notin \partial G(C) \ .$$

Case 3. If $-z_{(N+1)} < C < -z_{(N)}$, then $-z_{(j)} < C < -z_{(k)}$ for $j = N+1, \ldots, T,\ k = 1, \ldots, N$ so

$$0 = 1 - \frac{N}{N} = \partial G(C)\ .$$

Case 4. If $C = -z_{(N-p)} = -z_{(N+q)}$ with $-z_{(N+(q+1))} < C < -z_{(N-(p+1))}$ and $p \in \{0, 1, \ldots N-1\},\ q \in \{0, 1, \ldots, T-N\}$, then

$$0 \in \left[1 - \frac{N+q}{N},\ 1 - \frac{N-(p+1)}{N}\right] = \partial G(C)\ .$$

□

We now extend Lemma 3.3.2 to the cases where N in any non-integer real value in $[1, T]$.

Lemma 3.3.3. *If $N \in [1, T]$ is not an integer, then*

$$\arg\min_C G(C) = \{-z_{(\lfloor N \rfloor + 1)}\}\ .$$

Proof. Following the proof of Lemma 3.3.2, one shows that $0 \in \partial G(C)$ only when $C = -z_{(\lfloor N \rfloor + 1)}$ in which case

$$\partial G(C) = \left[1 - \frac{\lfloor N \rfloor + 1 + q}{N},\ 1 - \frac{\lfloor N \rfloor + 1 - p}{N}\right]$$

where $p \in \{1, \ldots, \lfloor N \rfloor\}$ and $q \in \{0, 1, \ldots, T - (\lfloor N \rfloor + 1)\}$ satisfy $-z_{(\lfloor N \rfloor + 1 + q)} = C = -z_{(\lfloor N \rfloor + 1 - p)}$. □

Summarizing Lemmas 3.3.2 and 3.3.3, the minimum of $G(C)$ is always attained at $C = -z_{(\lfloor N \rfloor + 1)}$ regardless of whether N integer or not. Choosing $N = \lfloor \alpha T \rfloor$, yields the following lemma.

Lemma 3.3.4. *Let $\frac{1}{T} \leq \alpha \leq 1$ and set $N = \lfloor \alpha T \rfloor$. Then*

$$\min_C G(C) = -\frac{1}{N} \sum_{t=1}^{N} z_{(t)} \tag{3.3.37}$$

where $z_{(t)}$ is the tth smallest element of z_t, $t = 1, \ldots, T$.

Proof. By Lemma 3.3.2, $\arg\min_C G(C) = [-z_{(N+1)}, -z_{(N)}]$. Let $C^* \in [-z_{(N+1)}, -z_{(N)}]$, then the right hand side of (3.3.37) becomes

$$\begin{aligned}
-\frac{1}{N}\sum_{t=1}^{N} z_{(t)} &= C^* + \frac{1}{N}\sum_{t=1}^{N}(-z_{(t)} - C^*) && (3.3.38)\\
&= C^* + \frac{1}{N}\sum_{t=1}^{N}[-z_{(t)} - C^*]_+ + \frac{1}{N}\sum_{t=N+1}^{T}[-z_{(t)} - C^*]_+ && (3.3.39)\\
&= C^* + \frac{1}{N}\sum_{t=1}^{T}[-z_{(t)} - C^*]_+, && (3.3.40)\\
&= G(C^*)\\
&= \min_C G(C)\ . && (3.3.41)
\end{aligned}$$

$\square$

Theorem 3.3.5. *If $N := \alpha T$ is an integer, $(\mathcal{P}_1)$ and $(\mathcal{P}_2)$ are exactly the same optimization problem.*

Proof. Set $z_t = w^\mathsf{T} r^t,\ t = 1, \ldots, T$. By Lemma 3.3.4, we have

$$\begin{aligned}
\min_{w,C}\ C + \frac{1}{\alpha T}\sum_{t=1}^{T}[-z_t - C]_+ \quad &= \quad \min_{w} \quad \left[\min_{C}\ C + \frac{1}{\alpha T}\sum_{t=1}^{T}[-z_t - C]_+\right]\\
\text{s.t. } w^\mathsf{T}\mu = \rho,\ w^\mathsf{T} e = 1 \qquad\qquad &\qquad \text{s.t. } w^\mathsf{T}\mu = \rho,\\
&\qquad\qquad w^\mathsf{T} e = 1\\[1em]
&= \quad \min_{w} \quad -\frac{1}{N}\sum_{t=1}^{N} z_{(t)} \qquad .\\
&\qquad \text{s.t. } w^\mathsf{T}\mu = \rho,\\
&\qquad\qquad w^\mathsf{T} e = 1
\end{aligned}$$

$\square$

3.4 ETG-ETL Portfolio Optimization

As discussed in Section 2.2, the mean is not a robust measure of a portfolio's reward. One improvement is to use ETG which takes the conditional expectation when the return is greater than the $1 - \beta$ quantile, where β is the probability of the right tail. It eliminates the influence of extreme negative returns, hence is considered a more robust measure of the reward. In this chapter, Section 3.4.1 formulates the ETG-ETL portfolio optimization problem; Chapter 4.1 develops an algorithm to solve this problem; a numerical experiment using real data from Yahoo Finance is presented.

3.4.1 ETG-ETL optimization problems

Since ETG is defined symmetrically to ETL, for $\beta \in (0, 1)$, we have

$$\mathrm{ETG}_\beta(r) = \mathrm{ETL}_\beta(-r). \tag{3.4.1}$$

In order to see this, let r be the return of an risky asset, $-r$ is then taken as the loss of the asset. Let $F_1(r)$ denote the cumulative distribution function of r, and $F_2(-r)$ denote the cumulative distribution function of $-r$, then for $\beta \in (0, 1)$, F_1 ad F_2 are related as follows:

$$F_1^{-1}(1 - \beta) = -F_2^{-1}(\beta). \tag{3.4.2}$$

Hence

$$\begin{align}
\mathrm{ETG}_\beta(r) &= E[r | r \geq F_1^{-1}(1 - \beta)] \tag{3.4.3}\\
&= -E[-r | -r \leq -F_1^{-1}(1 - \beta)] \tag{3.4.4}\\
&= -E[-r | -r \leq F_2^{-1}(\beta)] \tag{3.4.5}\\
&= \mathrm{ETL}_\beta(-r). \tag{3.4.6}
\end{align}$$

In the above proof, the first equality is the definition of ETG, the second equality plays a trick of flipping the sign of both sides of the inequality, the third equality uses equation (3.4.2) and the last equality is based on the definition of ETL. Combining this result (3.4.1) and formula (3.3.6) in Theorem 3.3.1, the ETG of a portfolio can also be written as

$$\begin{aligned}\mathrm{ETG}_\beta(w) &= \min_{D\in\mathbb{R}} \; H_\beta(w,D)\\ &= \min_{D\in\mathbb{R}} \; D + \frac{1}{\beta}\int [r_p - D]_+ p(r)dr,\end{aligned}$$

where $r_p = w^\mathsf{T} r$, the portfolio return, $p(r)$ is the probability density function for the portfolio's return r.

Proposition 3.4.1. *Given $\beta \in [0,1]$, the β-ETG of a portfolio, denoted by $ETG_\beta(w)$, is a convex function in w.*

Proof. Note that $H_\beta(w,D) = F_\beta(-w,D)$, where F is defined in (3.3.5). Since F is convex in D, H is also convex in D. Let D_i be such that $\mathrm{ETG}_\beta(w_i) = H_\beta(w_i, D_i)$, $i = 1,2$. For $\lambda \in [0,1]$, we have

$$\begin{aligned}
&\mathrm{ETG}_\beta(\lambda w_1 + (1-\lambda)w_2) && (3.4.7)\\
=&\min_{D\in\mathbb{R}} H_\beta(\alpha w_1 + (1-\alpha)w_2, D) && (3.4.8)\\
=&\min_{D\in\mathbb{R}} \alpha D + (1-\alpha)D + \frac{1}{\beta}\int [\alpha w_1^\mathsf{T} r + (1-\alpha)w_2^\mathsf{T} r && (3.4.9)\\
&- \alpha D - (1-\alpha)D]_+ p(r)dr, && (3.4.10)\\
\le&\,\alpha D_1 + (1-\alpha)D_2 + \frac{\alpha}{\beta}\int [w_1^\mathsf{T} r - D_1]_+ p(r)dr && (3.4.11)\\
&+ \frac{1-\alpha}{\beta}\int [w_2^\mathsf{T} r - D_2]_+ p(r)dr, && (3.4.12)\\
=&\,\alpha H_\beta(w_1, D_1) + (1-\alpha)H_\beta(w_2, D_2) && (3.4.13)\\
=&\,\alpha \mathrm{ETG}_\beta(w_1) + (1-\alpha)\mathrm{ETG}_\beta(w_2) && (3.4.14)
\end{aligned}$$

The second equality is just the definition of $H_\beta(w,D)$, the inequality is because $[\cdot]_+$ is a convex function and the last equality is true by the definition of D_i's. $\square$

A similar proof of the convexity of ETL is given in [31].

Proposition 3.4.2. *Given $\alpha \in [0,1]$, the α-ETL of a portfolio, denoted by $ETL_\alpha(w)$, is a convex function in w.*

Therefore both ETG and ETL of a portfolio are convex functions of w. The ETG-ETL portfolio optimization problem can be formulated either as a reward maximization problem with upper bound on risk or as a risk minimization problem with lower bound on reward. The mathematical expressions are shown below,

$$\begin{aligned} \underset{w\in\mathbb{R}}{\text{maximize}} \quad & \mathrm{ETG}_\beta(w) && (\mathcal{P}_3(\sigma_p)) \\ \text{subject to} \quad & \mathrm{ETL}_\alpha(w) \le \sigma_p && (3.4.15) \\ & w^\mathsf{T} e = 1 && (3.4.16) \end{aligned}$$

$$\begin{aligned} \underset{w\in\mathbb{R}}{\text{minimize}} \quad & \mathrm{ETL}_\alpha(w) && (\mathcal{P}_4(\mu_p)) \\ \text{subject to} \quad & \mathrm{ETG}_\beta(w) \ge \mu_p && (3.4.17) \\ & w^\mathsf{T} e = 1 && (3.4.18) \end{aligned}$$

where $\mathrm{ETG}_\beta(w) = \min_D H_\beta(w, D)$ and $\mathrm{ETL}_\alpha(w) = \min_C F_\alpha(w, C)$ are two convex functions, σ_p and μ_p are predetermined limit values for risk and reward.

Lemma 3.4.3. *In problem $\mathcal{P}_3(\sigma_p)$, the constraint* (3.4.15) *is active at every optimal solution, that is, for any optimal solution w^**

$$ETL_\alpha(w^*) = \sigma_p \ . \tag{3.4.19}$$

Proof. The result is clearly true if $\sigma_p \le \inf\{\mathrm{ETL}_\alpha(w) \,:\, w^\mathsf{T} e = 1\}$, so we assume that $\sigma_p > \inf\{\mathrm{ETL}_\alpha(w) \,:\, w^\mathsf{T} e = 1\}$. Denote the level set defined by (3.4.15) by

$$W = \{w |\ \mathrm{ETL}_\alpha(w) \le \sigma_p\}. \tag{3.4.20}$$

Since $\mathrm{ETL}_\alpha(w)$ is convex, W is also convex. Moreover, by [39, Theorem 7.6], the relative interior of W is

$$ri(W) = \{w |\ \mathrm{ETL}_\alpha(w) < \sigma_p\} \ . \tag{3.4.21}$$

Denote the feasible set of $\mathcal{P}_3(\sigma_p)$ by $\Omega(\sigma_p)$, i.e.

$$\Omega(\sigma_p) = \{w|\ \mathrm{ETL}_\alpha(w) \leq \sigma_p,\ w^\mathsf{T} e = 1\}. \tag{3.4.22}$$

Since it is the intersection of a convex set and an affine set, it is still a convex set. Also define

$$\Omega_0(\sigma_p) = \{w|\ \mathrm{ETL}_\alpha(w) < \sigma_p,\ w^\mathsf{T} e = 1\}\ . \tag{3.4.23}$$

We claim that $\Omega_0(\sigma_p) \subset ri(\Omega(\sigma_p))$. Let $\hat{w} \in \Omega_0(\sigma_p)$, we need to show that for every $w \in \Omega(\sigma_p)$, $\exists\ \lambda > 1$, s.t. $(1-\lambda)w + \lambda\hat{w} \in \Omega(\sigma_p)$ [39, Theorem 6.4]. Let $w \in \Omega(\sigma_p) \subset W$. Since $\Omega_0(\sigma_p) \subset ri(W)$, then $\exists\ \lambda > 1$ such that

$$(1-\lambda)w + \lambda\hat{w} \in W \tag{3.4.24}$$

we also have

$$((1-\lambda)w + \lambda\hat{w})^\mathsf{T} e = (1-\lambda)w^\mathsf{T} e + \lambda\hat{w}^\mathsf{T} e = 1 \tag{3.4.25}$$

So $(1-\lambda)w + \lambda\hat{w} \in \Omega(\sigma_p)$. Hence $\hat{w} \in ri(\Omega(\sigma_p))$ establishing the claim.

By [39, Corollary 32.2.1], every optimal solution to $(\mathcal{P}_3(\sigma_p))$ must lie on the relative boundary of $\Omega(\sigma_p)$. Hence the constraint (3.4.15) is active at all solutions since $\Omega_0(\sigma_p) \subset ri(\Omega(\sigma_p))$. □

Theorem 3.4.4. *Let α and β be fixed values in $[0,1]$. Let $V_3(\sigma_p)$ be the optimal value function for $\mathcal{P}_3(\sigma_p)$ as a function of σ_p, and similarly let $V_4(\mu_p)$ be the optimal value function for $\mathcal{P}_4(\mu_p)$ as a function of μ_p.*

(i) If w^* solves $\mathcal{P}_3(\sigma_p)$, then w^* solves $\mathcal{P}_4(V_3(\sigma_p))$ and

$$\sigma_p = V_4(V_3(\sigma_p))\ .$$

(ii) If w^* solves $\mathcal{P}_4(\mu_p)$ with $ETG_\beta(w^*) = \mu_p$, then w^* solves $\mathcal{P}_3(V_4(\mu_p))$ and

$$\mu_p = V_3(V_4(\mu_p))\ .$$

Proof. We only prove (i) since (ii) is proved in the same way. Define $\mu_p = V_3(\sigma_p) = \mathrm{ETG}_\beta(w^*)$. Then w^* is feasible for $\mathcal{P}_4(\mu_p)$, and so

$$V_4(\mu_p) \leq \mathrm{ETL}_\alpha(w^*) = \sigma_p$$

where the final equality follows from Lemma 3.4.3. If w is feasible for $\mathcal{P}_4(\mu_p)$, then $\mathrm{ETG}_\beta(w) \geq \mu_p = V_3(\sigma_p) = \mathrm{ETG}_\beta(w^*)$ and so we must have

$$\mathrm{ETL}_\alpha(w) \geq \sigma_p = \mathrm{ETL}_\alpha(w^*) \geq V_4(\mu_p) \tag{3.4.26}$$

since μ_p is the optimal value for $\mathcal{P}_3(\sigma_p)$. Hence w^* solves $\mathcal{P}_4(\mu_p)$, and by minimizing the left-hand side of (3.4.26) over all w feasible for $\mathcal{P}_4(\mu_p)$, we have

$$\sigma_p = \mathrm{ETL}_\alpha(w^*) = V_4(\mu_p) = V_4(V_3(\sigma_p)).$$

$\square$

Chapter 4

AN ALGORITHM FOR ETG-ETL PORTFOLIO OPTIMIZATION

In this chapter, an iterative algorithm is developed to solve the ETG-ETL optimization problem using empirical estimations $\widehat{\text{ETG}}$ and $\widehat{\text{ETL}}$. Notice that in the problem $\mathcal{P}_3(\sigma_p)$, the feasible set is the intersection of a hyperplane defined by $w^\mathsf{T} e = 1$ and the σ_p-*sublevel set* defined by $\text{ETL}_\alpha(w) \leq \sigma_p$, and thus is a convex set. However, the feasible set for the problem $\mathcal{P}_4(\mu_p)$ is not convex. For this reason, it is more convenient to work with the problme $\mathcal{P}_3(\sigma_p)$, which is maximizing a convex function over a convex set. Since maximizing a convex function is equivalent to minimizing the negative of the objective function, a concave function, this is a concave programming problem. We replace ETG and ETL with corresponding empirical estimations in $\mathcal{P}_3(\sigma_p)$, to obtain the problem,

$$\begin{aligned} \underset{w\in\mathbb{R}}{\text{maximize}} \quad & \widehat{\text{ETG}}_\beta(w) \\ \text{subject to} \quad & \widehat{\text{ETL}}_\alpha(w) \leq \sigma_p \\ & w^\mathsf{T} e = 1, \end{aligned} \tag{$\mathcal{P}_3'$}$$

where the functions in $\mathcal{P}_3'$ are empirical expected tail gain and expected tail loss defined by

$$\widehat{\text{ETG}}_\beta(w) \;:=\; \min_{D\in\mathbb{R}} D + \frac{1}{\beta T}\sum_{t=1}^{T}[w^\mathsf{T} r^t - D]_+ \tag{4.0.1}$$

$$\widehat{\text{ETL}}_\alpha(w) \;:=\; \min_{C\in\mathbb{R}} C + \frac{1}{\alpha T}\sum_{t=1}^{T}[-w^\mathsf{T} r^t - C]_+ \leq \sigma_p\ . \tag{4.0.2}$$

As in Lemma 3.4.3, the optimal solution to this problem lies on the relative boundary of the feasible set. Briefly, the proposed algorithm proceeds as follows. The algorithm is initiated at a point in the relative interior of the feasible set, and then searches along an initial set of directions spanning the affine hull of the feasible set. For each direction, it first follows the maximum ascent direction of $\widehat{\text{ETG}}$ to reach the relative boundary of the feasilbe set, then searches along the boundary until it reaches a local maximum. Finally it reports the maximum of all the local maximum values as the current best estimate of the optimal value.

4.1 Support function expressions for $\widehat{ETG}$ and $\widehat{ETL}$

Since both $\widehat{\text{ETG}}$ and $\widehat{\text{ETL}}$ require minimization to obtain there values, problem $\mathcal{P}_3'$ is difficult to handle directly. We begin by obtaining alternative representations of these functions as support functions.

Definition 4.1.1. The support function $h_A : \mathbb{R}^n \to \mathbb{R}$ of a non-empty closed convex set A in $\mathbb{R}^n$ is given by

$$h_A(x) := \sup_{a \in A} < x, a >, \tag{4.1.1}$$

where $< x, a >$ is the inner product of x and a. In the Euclidian space as discussed in this disseration, $< x, a >= x^{\mathsf{T}} a$. These two notations are used interchangablely in this context.

A support function is a convex function on $\mathbb{R}^n$. Conversely, any convex positive homogeneous function (or sublinear function) on $\mathbb{R}^n$ is the support function of a non-empty convex set. The underlying convex set can be taken to be compact if and only if the support function is everywhere finite-valued. See [41] for the proof.

Definition 4.1.2. A *sublinear* function $f(x) : \mathbb{R}^n \to \mathbb{R}$ is a function satisfying positive homogeneity and subadditivity, i.e.

$$f(\gamma x) = \gamma f(x) \text{ for any } \gamma > 0 \text{ and } x \in \mathbb{R}^n$$

$$f(x + y) \leq f(x) + f(y) \text{ for any } x, y \in \mathbb{R}^n.$$

ETL is a coherent risk measure, so it is sublinear. Equation (3.4.1) shows that ETG is sublinear as well. The empirical estimates $\widehat{\text{ETG}}$ and $\widehat{\text{ETL}}$ are positive are also sublinear. The fact that they are positive homogeneous is easily proved by definition. And the fact that $\widehat{\text{ETG}}$ is convex can be shown by replacing the integral in ETL by its empirical estimation in the proof of Proposition (3.4.1). A similar proof shows that $\widehat{\text{ETL}}$ is also convex. Hence, since both of the empirical estimations are everywhere finite-valued, they can be represented as the support functions of some nonempty compact convex sets. We now derive these support function representations. For $\widehat{\text{ETL}}$, we have

$$
\begin{aligned}
\widehat{\text{ETL}}_\alpha(w) &= \min_C C + \frac{1}{\alpha T}\sum_{t=1}^{T}[-w^\mathsf{T} r^t - C]_+ && (4.1.2)\\
&= \min_C C + \frac{1}{\alpha T}\sum_{t=1}^{T}\max\{-w^\mathsf{T} r^t - C, 0\} && (4.1.3)\\
&= \min_C C + \sum_{t=1}^{T}\sup_{0\le p_t\le 1/\alpha T} p_t(-w^\mathsf{T} r^t - C) && (4.1.4)\\
&= \min_C C + \sup_{0\le p\le \frac{1}{\alpha T}e} < p, -Rw - Ce > && (4.1.5)\\
&= \min_C \max_{0\le p\le \frac{1}{\alpha T}e} < p, -Rw > + C(1- < p, e >) && (4.1.6)\\
&\ge \max_{0\le p\le \frac{1}{\alpha T}e} \min_C < p, -Rw > + C(1- < p, e >) && (4.1.7)\\
&= \max_{\substack{-\frac{1}{\alpha T}e\le p\le 0\\ <p,e>=-1}} < p, Rw > . && (4.1.8)
\end{aligned}
$$

In equation (4.1.4), we introduce the auxiliary variables $p_t, t = 1, \ldots, T$, and in equation (4.1.5) we introduce the vector for these variables, $p = (p_1, \ldots, p_T)^\mathsf{T}$. We claim that equality holds in (4.1.7). To see this, flip the sign of p in (4.1.8) and rewrite it as the following problem

$$\underset{p}{\text{minimize}} \quad < p, Rw > \tag{$\mathcal{P}$}$$

$$\text{subject to} \quad 0 \leq p \leq \frac{1}{\alpha T} e \tag{4.1.9}$$

$$< p, e >= 1 \tag{4.1.10}$$

The Lagrangian for this linear program is

$$L(p; C, \lambda, \delta) =< p, Rw > +C(< p, e > -1) - \lambda^{\mathsf{T}} p + \delta^{\mathsf{T}}(p - \frac{1}{\alpha T} e), \tag{4.1.11}$$

where p is the primal variable and C, λ, δ are the dual variables. Since

$$\max_{C,\lambda \geq 0, \delta \geq 0} L(p; C) = \begin{cases} < p, Rw > & 0 \leq p \leq \frac{1}{\alpha T} e, \ < p, e >= 1 \\ +\infty, & otherwise, \end{cases} \tag{4.1.12}$$

the primal problem is

$$p^* = \min_p \max_{C,\lambda \geq 0, \delta \geq 0} \quad L(p; C), \tag{4.1.13}$$

or equivalently problem $(\mathcal{P})$. This is a linear programming problem, so strong duality holds since this LP is feasible with compact feasible region. The dual problem is easily obtained as follows:

$$d^* = \max_{C,\lambda \geq 0, \delta \geq 0} \min_p \quad L(p; C) \tag{4.1.14}$$

$$= \max_{C,\lambda \geq 0, \delta \geq 0} \min_p \quad p^{\mathsf{T}}(Rw + Ce + \sigma - \lambda) - C - \frac{1}{\alpha T} \sigma^{\mathsf{T}} e \tag{4.1.15}$$

$$= \max_{C,\lambda \geq 0, \delta \geq 0} -C - \frac{1}{\alpha T} \sigma^{\mathsf{T}} e \quad \text{subject to} \quad Rw + Ce + \sigma - \lambda = 0, \tag{4.1.16}$$

or equivalently,

$$\min_{\delta, C} \quad C + \frac{1}{\alpha T} \delta^{\mathsf{T}} e \tag{$\mathcal{D}$}$$

$$\text{subject to} \quad \delta \geq -Rw - Ce \tag{4.1.17}$$

$$\delta \geq 0. \tag{4.1.18}$$

Observe that this is exactly the optimization problem for $\widehat{\text{ETL}}$ as stated in Theorem 3.3.1. Given that strong duality holds, their optimal values p^* and d^* coincide, i.e. the equality in (4.1.7) holds. So we have,

Proposition 4.1.1. *The support function expression for $\widehat{ETL}$ is*

$$\widehat{ETL}_\alpha(w) = \max_{\substack{-\frac{1}{\alpha T}e \leq p \leq 0 \\ <p,e>=-1}} < p, Rw > . \tag{4.1.19}$$

In the same way, we can derive the support function expression for $\widehat{\text{ETG}}$.

Proposition 4.1.2. *The support function expression for $\widehat{ETG}$ is*

$$\widehat{ETG}_\alpha(w) = \max_{\substack{0 \leq q \leq \frac{1}{\alpha T}e \\ <q,e>=1}} < q, Rw > . \tag{4.1.20}$$

Notice that the sets $\{R^\mathsf{T} p : -\frac{1}{\alpha T}e \leq p \leq 0, < p, e >= -1\}$ and $\{R^\mathsf{T} q : 0 \leq q \leq \frac{1}{\alpha T}e, < q, e >= 1\}$ are the non-empty compact convex sets for the two support functions, respectively. In the next section, we derive the subdifferential of $\widehat{\text{ETG}}_\beta(w)$ from its support function expression (4.1.20).

4.2 Subdifferential of $\widehat{\text{ETG}}_\beta(w)$

In this section we derive a closed form expression for the subdifferential of $\widehat{\text{ETG}}_\beta(w)$ for fixed weights w. The key fact we use in this derivation is that for any support function h_A, as in definition 4.1.1, we have

$$\partial h_A(x) = \arg\max_{a \in A} \langle x, a \rangle$$

(e.g. see [39, Corollary 23.5.3]).

Recall that the primal and dual optimization problems for $\widehat{\text{ETG}}_\beta(w)$ are

$$\widehat{\text{ETG}}_\beta(w) = \min_D D + \frac{1}{\beta T} \sum_{t=1}^{T} [w^\mathsf{T} r^t - D]_+ \tag{4.2.1}$$

$$= \max_{\substack{0 \leq q \leq \frac{1}{\beta T}e \\ q^\mathsf{T} e = 1}} q^\mathsf{T} Rw. \tag{4.2.2}$$

Set $\tilde{w} = Rw$, then when w is fixed, $\tilde{w}$ is fixed as well. Assume βT is an integer and denote it by K $(0 < K \leq T)$, then $\widehat{\mathrm{ETG}}_\beta(w)$ can be expressed as

$$\max q^\mathsf{T}\tilde{w}, \quad \text{subject to} \quad 0 \leq q \leq \frac{1}{K}e, \quad q^\mathsf{T}e = 1 \tag{4.2.3}$$

It is obvious that its optimal solution q^* is obtained by picking the K largest elements of $\tilde{w}$. To make this rigrous, we introduce some notation. Let $\tilde{w}_{l(k)}$ be the kth largest element of $\tilde{w}$ counting multiplicities, and define index sets $I^>(\tilde{w}, k)$, $I^=(\tilde{w}, k)$ and $I^\geq(\tilde{w}, k)$ as follows

$$I^>(\tilde{w}, k) = \{i : \tilde{w}_i > \tilde{w}_{l(k)}\} \tag{4.2.4}$$

$$I^=(\tilde{w}, k) = \{i : \tilde{w}_i = \tilde{w}_{l(k)}\} \tag{4.2.5}$$

$$I^<(\tilde{w}, k) = \{i : \tilde{w}_i < \tilde{w}_{l(k)}\}. \tag{4.2.6}$$

Accordingly, $I^\geq(\tilde{w}, k) = I^>(\tilde{w}, k) \cup I^=(\tilde{w}, k)$. Also let $N^>(k)$, $N^=(k)$ and $N^<(k)$ represent the cardinality of corresponding sets, respectively. Figure 4.1 gives two examples when the cardinality of $I^=(\tilde{w}, k)$ is one, i.e. $N^=(k) = 1$ and when it is greater than 1, i.e. $N^=(k) > 1$. Then the optimal solution set $\{q^*\}$ is given by

$$Q(\tilde{w}, K) :=$$

$$\{q : q_i = \frac{1}{K}, \quad \text{for } i \in I^>(\tilde{w}, K); \tag{4.2.7}$$

$$q_i = \tfrac{K - N^>(K)}{K}\mu_i, \quad \sum_{i \in I^=(\tilde{w}, K)} \mu_i = 1, 0 \leq \mu_i \leq \tfrac{1}{K - N^>(K)} \text{ for } i \in I^=(\tilde{w}, K); \tag{4.2.8}$$

$$q_i = 0, \quad \text{for } i \in I^<(\tilde{w}, K)\} \tag{4.2.9}$$

$$= \{q : q_i = \frac{1}{K}, \quad \text{for } i \in I^>(\tilde{w}, K); \tag{4.2.10}$$

$$q_i = \frac{1}{K}\omega_i, \quad \sum_{i \in I^=(\tilde{w}, K)} \omega_i = [K - N^>(K)], 0 \leq \omega_i \leq 1 \text{ for } i \in I^=(\tilde{w}, K); \tag{4.2.11}$$

$$q_i = 0, \quad \text{for } i \in I^<(\tilde{w}, K)\}. \tag{4.2.12}$$

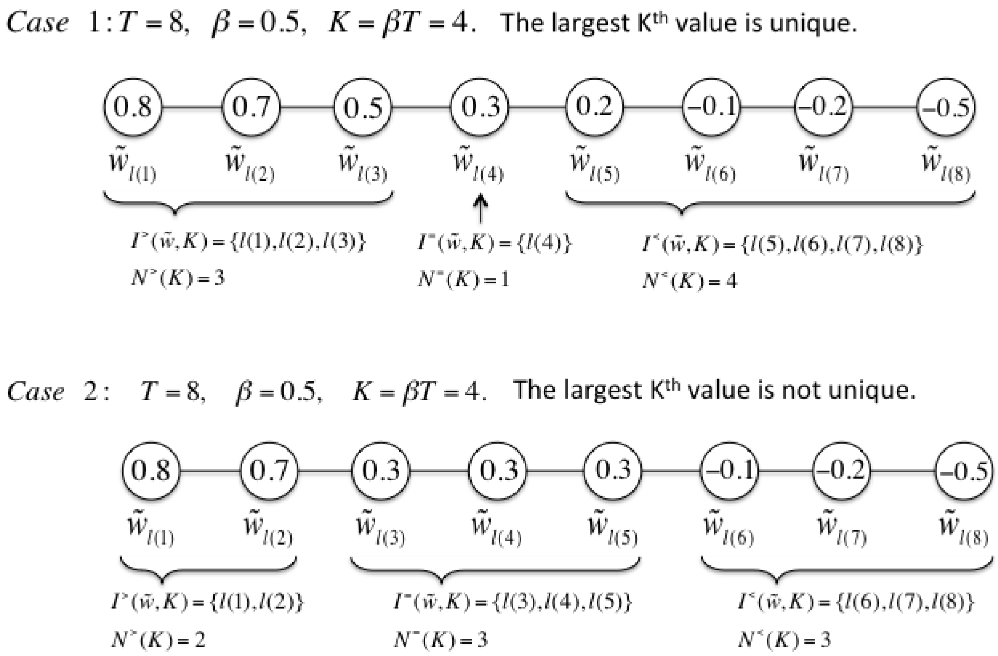

Figure 4.1: Illustration of $I^{>}(\tilde{w}, k)$, $I^{=}(\tilde{w}, k)$ and $I^{\geq}(\tilde{w}, k)$

Notice that in (4.2.8), the μ_i are the coefficients in a convex combination of the elements of $\tilde{w}_i$ with $i \in I^{=}(\tilde{w}, K)$, but not all possible convex combinations are allowed. If the multiplicity of the K^{th} element is 1, there is only one optimal solution q^*, which assigns weight $\frac{1}{K}$ to each of the largest K elements of $\tilde{w}$. If the multiplicity of the K^{th} element is strictly greater than 1, there are an infinite number of solutions for q^*, each having an assigned weight of $\frac{1}{K}$ to the elements with index in $I^{>}(\tilde{w}, K)$, a weight of $\frac{K-N^{>}(K)}{K}\mu_i$ to each i in $I^{=}(\tilde{w}, K)$ with $\sum_{i \in I^{=}(\tilde{w},K)} \mu_i = 1$, $0 \leq \mu_i \leq \frac{1}{K-N^{>}(K)}$, and a weight of 0 to the remaining indices. We also make use of the following relaxation of the set $Q(\tilde{w}, K)$:

$$
\begin{aligned}
&Q_r(\tilde{w}, K) := \\
&\qquad \{q : q_i = \frac{1}{K}, \quad \text{for } i \in I^{>}(\tilde{w}, K); \\
&\qquad\quad q_i = \tfrac{K - N^{>}(K)}{K} \mu_i, \quad \sum \mu_i = 1, 0 \le \mu_i \text{ for } i \in I^{=}(\tilde{w}, K); \\
&\qquad\quad q_i = 0, \quad \text{for } i \in I^{<}(\tilde{w}, K)\}.
\end{aligned} \tag{4.2.13}
$$

These sets only differ in defining the expressions for $\mu_i, i \in I^{=}(\tilde{w}, K)$. In $Q_r(\tilde{w}, K)$ we allow all possible convex combinations of $\mu_i, i \in I^{=}(\tilde{w}, K)$.

Based on this representation for an optimal solution q^*, we have the following closed form expression for the subdifferential of $\widehat{\text{ETG}}_\beta(w)$:

$$
\begin{aligned}
\partial \widehat{\text{ETG}}_\beta(w) &= \{R^\mathsf{T} q^* | q^* \in Q(\tilde{w}, K)\} && (4.2.14) \\
&\subset \{R^\mathsf{T} q^* | q^* \in Q_r(\tilde{w}, K)\} && (4.2.15) \\
&= \frac{1}{K} \sum_{i \in I^{>}(\tilde{w}, K)} R_{i\cdot}^\mathsf{T} + \frac{K - N^{>}(K)}{K} \text{Conv}\left\{R_{i\cdot}^\mathsf{T} | i \in I^{=}(\tilde{w}, K)\right\} && (4.2.16) \\
&=: \partial_r \widehat{\text{ETG}}_\beta(w) \ , && (4.2.17)
\end{aligned}
$$

where $R_{i\cdot}$ is the i^{th} row of R and Conv$\{A\}$ represents the convex hull of a set A. We call the final expression $\partial_r \widehat{\text{ETG}}_\beta(w)$ the relaxed subdifferential for $\widehat{\text{ETG}}_\beta$ at w. If the cardinality of $I^{=}(\tilde{w}, K)$ is 1, i.e., $N^{=}(K) = 1$, there is only one μ with value 1 and $\widehat{\text{ETG}}_\beta(w)$ is differentiable and its derivative is

$$
\widehat{\text{ETG}}'_\beta(w) = \frac{1}{K} \sum_{i \in I^{\ge}(\tilde{w}, K)} R_{i\cdot}^\mathsf{T} \tag{4.2.18}
$$

To summarize, we have proved the following claim.

Claim 1. The subdifferential of $\widehat{\text{ETG}}_\beta(w)$ is

$$
\partial \widehat{\text{ETG}}_\beta(w) = \{R^\mathsf{T} q^* | q^* \in Q(\tilde{w}, K)\}. \tag{4.2.19}
$$

Note that we always have

$$q(\tilde{w}, K) = \frac{1}{K}\left[\sum_{i\in I^{>}(\tilde{w},K)} R_{i\cdot}^{\mathsf{T}} + \frac{K - N^{>}(K)}{N^{=}(K)} \sum_{i\in I^{=}(\tilde{w},K)} R_{i\cdot}^{\mathsf{T}}\right] \in \partial\widehat{\mathrm{ETG}}_{\beta}(w). \qquad (4.2.20)$$

In the next two sections we find the analytical formulas for directions of ascent for $\widehat{\mathrm{ETG}}_{\beta}(w)$ in two cases: when w is in the relative interior of the feasible set and when it is on the relative boundary of the feasible set.

4.3 Directions of ascent for $\widehat{\mathrm{ETG}}_{\beta}(w)$ when w is in the relative interior

Given w fixed in the relative interior of the feasible set, we seek a direction Δw that gives a positive directional derivative, or, if possible, a direction of steepest ascent for $\widehat{\mathrm{ETG}}_{\beta}(w)$. Denote the directional derivative of $\widehat{\mathrm{ETG}}_{\beta}(w)$ at w in the direction Δw by $\widehat{\mathrm{ETG}}_{\beta}'(w; \Delta w)$:

$$\widehat{\mathrm{ETG}}_{\beta}'(w; s) := \lim_{\tau\downarrow 0} \frac{\widehat{\mathrm{ETG}}_{\beta}(w + \tau s) - \widehat{\mathrm{ETG}}_{\beta}(w)}{\tau} .$$

Since $\widehat{\mathrm{ETG}}_{\beta}$ is convex, we have, by [39, Theorem 23.4], that

$$\widehat{\mathrm{ETG}}_{\beta}'(w; s) = \max_{z\in\partial\widehat{\mathrm{ETG}}_{\beta}(w)} \langle z,\, s\rangle . \qquad (4.3.1)$$

We are interested in the direction of steepest ascent for $\widehat{\mathrm{ETG}}_{\beta}(w)$ relative to the manifold $e^{\mathsf{T}}w = 1$. This is given as the is the solution to the optimization problem

$$\begin{aligned} &\underset{s\in\mathbb{R}^n}{\text{maximize}} \quad \widehat{\mathrm{ETG}}_{\beta}'(w; s) && (\mathcal{P}_{a1})\\ &\text{subject to} \quad s^{\mathsf{T}}e = 0,\ \|s\|_2 \le 1 . && (4.3.2)\end{aligned}$$

The constraint $s^{\mathsf{T}}e = 0$ assures that the direction s is tangential to the manifold $e^{\mathsf{T}}w = 1$. Indeed, if we start from a feasible solution w satisfying $w^{\mathsf{T}}e = 1$, we want to find a search direction Δw such that $w + t\Delta w$ still satisfies $(w + t\Delta w)^{\mathsf{T}}e = 1$. For the second constraint, we use 2-norm to confine the length of the direction vector because this enable us to get a closed form solution to an approximation to this problem. The details are given below.

By (4.3.1), we can rewrite $(\mathcal{P}_{a1})$ as

$$\underset{s,q\in\mathbb{R}^n}{\text{maximize}} \quad q^\mathsf{T} s \tag{4.3.3}$$

$$\text{subject to} \quad s^\mathsf{T} e = 0,\ \|s\|_2 \le 1,\ q \in \partial\widehat{\mathrm{ETG}}_\beta(w). \tag{4.3.4}$$

For ease of computation, we relax this problem to

$$\underset{s,q\in\mathbb{R}^n}{\text{maximize}} \quad q^\mathsf{T} s \tag{4.3.5}$$

$$\text{subject to} \quad s^\mathsf{T} e = 0,\ \|s\|_2 \le 1,\ q \in \partial_r\widehat{\mathrm{ETG}}_\beta(w). \tag{4.3.6}$$

Using the definition of $\partial_r\widehat{\mathrm{ETG}}_\beta(w)$, we can write this problem as

$$\frac{1}{K}\max_{\substack{s^\mathsf{T} e=0\\ \|s\|_2\le 1}}\left[\sum_{i\in I^>(\tilde{w},K)} R_{i\cdot}s + (K-N^>(K))\max_{\substack{\sum\mu_i=1,\ \mu_i\ge 0\\ i\in I^=(\tilde{w},K)}}\sum \mu_i R_{i\cdot}s\right] \tag{4.3.7}$$

$$=\frac{1}{K}\max_{\substack{s^\mathsf{T} e=0\\ \|s\|_2\le 1}}\left[\sum_{i\in I^>(\tilde{w},K)} R_{i\cdot}s + (K-N^>(K))\max_{j\in I^=(\tilde{w},K)} R_{j\cdot}s\right] \tag{4.3.8}$$

$$=\frac{1}{K}\max_{j\in I^=(\tilde{w},K)}\max_{\substack{s^\mathsf{T} e=0\\ \|s\|_2\le 1}}\left[\sum_{i\in I^>(\tilde{w},K)} R_{i\cdot} + (K-N^>(K))R_{j\cdot}\right]s. \tag{4.3.9}$$

From (4.3.7) to (4.3.8), we observe that the maximum is attained when one of the μ_i's is 1, and 0 for the others. It gives the maximum value of $R_{i\cdot}s$. For each $j \in I^=(\tilde{w}, K)$, define a vector $e^{(j,w)}$ as

$$e_i^{(j,w)} = \begin{cases} 1, & i \in I^>(\tilde{w},K); \\ K - N^>(K), & i = j; \\ 0, & \text{otherwise.} \end{cases} \tag{4.3.10}$$

Also let $\hat{q}_{(j,w)} = R^\mathsf{T} e^{(j,w)}$, then (4.3.9) can be written as

$$\max_{j\in I^=(\tilde{w},K)}\max_{\substack{s^\mathsf{T} e=0\\ \|s\|_2\le 1}} \hat{q}_{(j,w)}^\mathsf{T} s. \tag{4.3.11}$$

We dropped the factor $\frac{1}{K}$ in the above problem, since it makes no difference for the optimal solution. For the inner maximization problem, there is a closed form solution. We state it as a lemma.

Lemma 4.3.1. *Consider the problem*

$$\begin{aligned} \max_{s\in\mathbb{R}^n} \quad & v^{\mathsf{T}} s \\ \textit{subject to} \quad & e^{\mathsf{T}} s = 0, \quad \|s\|_2 \le 1 \end{aligned} \tag{$\mathcal{P}_5$}$$

Let the operator $P = I - \frac{ee^{\mathsf{T}}}{n}$*, then* P *is an orthogonal projector onto the null space of* e^{T}*. Also*

$$\bar{s} = \frac{Pv}{\|Pv\|_2}$$

gives the optimal solution to ($\mathcal{P}_5$*) and the optimal value is* $\|Pv\|_2$*.*

Proof. 1. We prove the first statement in three steps.

(i) $P^2 = P$, so P is a projector.

$$\begin{aligned} P^2 &= \left[I - \frac{ee^{\mathsf{T}}}{n}\right]\left[I - \frac{ee^{\mathsf{T}}}{n}\right] \\ &= I - 2\frac{ee^{\mathsf{T}}}{n} + \frac{ee^{\mathsf{T}}}{n} \\ &= I - \frac{ee^{\mathsf{T}}}{n} \\ &= P \end{aligned}$$

(ii) It is easy to check that $P = P^{\mathsf{T}}$, so P is an orthognal projector.

(iii) It is also easy to check that $Pe = 0$. Since rank$(P) = n - 1$, so e forms a base of the null space of P, and null(e^{T}) =range(P).

Summarizing the above three steps, we show that P is an orthogonal projector onto the null space of e^{T}.

2. Since P is an orthogonal projector onto $\mathrm{null}(e^\mathsf{T})$, we can reformulate the problem $(\mathcal{P}_5)$ as

$$\begin{aligned} \max_{s\in\mathbb{R}^n} \quad & v^\mathsf{T} P s = (Pv)^\mathsf{T}(Ps) \\ \text{subject to} \quad & \|Ps\|_2 \le 1 \end{aligned} \tag{$\mathcal{P}_6$}$$

Replace Ps with z, we can further rewrite it as

$$\begin{aligned} \max_{z} \quad & (Pv)^\mathsf{T} z \\ \text{subject to} \quad & \|z\|_2 \le 1, \quad z \in \mathrm{range}(P) \end{aligned} \tag{$\mathcal{P}_7$}$$

Obviously, without the constraint $z \in \mathrm{range}(P)$, the optimal solution is $z^* = \frac{Pv}{\|Pv\|_2}$. Since $Pv \in \mathrm{range}(P)$, z^* is also the optimal solution to $(\mathcal{P}_7)$. Moreover, $Pz^* = \frac{P^2v}{\|Pv\|_2} = z^*$, so z^* is also the optimal solution to $(\mathcal{P}_6)$. Since $(\mathcal{P}_5)$ and $(\mathcal{P}_6)$ are the same problem, the optimal solution to $(\mathcal{P}_5$ is

$$s^* = \frac{Pv}{\|Pv\|_2}, \tag{4.3.12}$$

and its optimal value is

$$v^\mathsf{T} s^* = \frac{(Pv)^\mathsf{T}(Pv)}{\|Pv\|_2} = \|Pv\|_2 \,. \tag{4.3.13}$$

□

Based on the Lemma 4.3.1, the relaxed direction of steepest ascent is obtained by solving the problem

$$\max_{j\in I^=(\tilde{w},K)} \left\|P\hat{q}^{(j,w)}\right\|_2 \tag{$\mathcal{P}'_{ascent}$}$$

where $P = I - ee^\mathsf{T}/n$ and $\hat{q}^{(j,w)} = R^\mathsf{T} e^{(j,w)}$, $e^{(j,w)}$ is defined in (4.3.10). By using the direction $s = \frac{Pq(\tilde{w},K)}{\|Pq(\tilde{w},K)\|_2}$ with $q(\tilde{w},K)$ defined in (4.2.20), we have the following lower bound on the optimal value in $\mathcal{P}'_{ascent}$:

$$\begin{aligned} K\left\|Pq(\tilde{w},K)\right\|_2 &= \left\langle Kq(\tilde{w},K),\ \frac{Pq(\tilde{w},K)}{\|Pq(\tilde{w},K)\|_2}\right\rangle \\ &\leq \max_{z\in K\partial\widehat{\mathrm{ETG}}_\beta(w)} \left\langle z,\ \frac{Pq(\tilde{w},K)}{\|Pq(\tilde{w},K)\|_2}\right\rangle \\ &\leq \max_{z\in K\partial_r\widehat{\mathrm{ETG}}_\beta(w)} \left\langle z,\ \frac{Pq(\tilde{w},K)}{\|Pq(\tilde{w},K)\|_2}\right\rangle . \end{aligned}$$

This guarantees a positive value for the relaxed maximum ascent direction as long as $q(\tilde{w},K)$ is not in the null space of e^{T}, which is true in this case.

4.4 Steepest ascent direction of $\widehat{ETG}_\beta(w)$ when w is on the relative boundary

In this section, we add one more constraint to the previous problem $(\mathcal{P}_{a1})$,

$$\widehat{\mathrm{ETL}}'_\alpha(w;\Delta w) \leq 0. \tag{4.4.1}$$

This makes $\widehat{\mathrm{ETL}}_\alpha(w)$ remain or decrease when we move in the direction Δw for a sufficiently short distance.

Theorem 4.4.1. *The* σ_p-sublevel set *defined by* $\{w|\widehat{ETL}_\alpha(w) \leq \sigma_p\}$ *is a convex polyhedron.*

Proof. According to the support function expression for $\widehat{\mathrm{ETL}}$ in (4.1.19), the σ_p-sublevel set can also be represented as

$$\left\{ w \,\middle|\, \max_{\substack{-\frac{1}{\alpha T}e\leq p\leq 0\\ <p,e>=-1}} p^{\mathsf{T}}Rw \leq \sigma_p \right\} \tag{4.4.2}$$

$$= \bigcap_{\substack{-\frac{1}{\alpha T}e\leq p\leq 0\\ <p,e>=-1}} \{w|p^{\mathsf{T}}Rw \leq \sigma_p\}. \tag{4.4.3}$$

The set $\mathcal{T} := \{p : -\frac{1}{\alpha T}e \leq p \leq 0,\ <p,e>=-1\}$ is a convex polytope and so has a finite number of extreme points, which we denote as $\mathcal{E}$, with $\mathcal{T} = \mathrm{conv}(\mathcal{E})$. It is straightforward to show that

$$\bigcap_{p\in\mathcal{T}}\{w|p^{\mathsf{T}}Rw \le \sigma_p\} = \bigcap_{p\in\mathcal{E}}\{w|p^{\mathsf{T}}Rw \le \sigma_p\}.$$

Hence $\{w|\widehat{\mathrm{ETL}}_\alpha(w) \le \sigma_p\}$ can be represented as the intersection of finitely many half-spaces, that is, it is a convex polyhedron. □

Theorem 4.4.1 guarantees that for a sufficiently small t, $w + t\Delta w$ remains within the σ_p-sublevel $\{w|\widehat{\mathrm{ETL}}_\alpha(w) \le \sigma_p\}$. Now the problem is formulated as

$$\begin{aligned} \underset{s\in\mathbb{R}^n}{\text{maximize}} \quad & \widehat{\mathrm{ETG}}'_\beta(w;s) && (\mathcal{P}_{a2}) \\ \text{subject to} \quad & s^{\mathsf{T}}e = 0,\ \|s\|_\infty \le 1 && (4.4.4) \\ & \widehat{\mathrm{ETL}}'_\alpha(w;s) \le 0, && (4.4.5) \end{aligned}$$

which yields a direction of steepest descent for $\widehat{\mathrm{ETG}}_\beta$ tangential to the manifold

$$\Omega := \left\{ w \;\middle|\; w^{\mathsf{T}}e = 1,\ \widehat{\mathrm{ETL}}_\alpha(w) \le \sigma_p \right\} .$$

Note that the norm constraint is defined as infinity norm for the ease of computation, the reason is clear later. By paralleling the approach to the computation of $\partial\widehat{\mathrm{ETG}}_\beta(w)$, we derive an expression for the subdifferential of $\widehat{\mathrm{ETL}}_\alpha(w)$. Recall from Proposition 4.1.1 that the support function expression for $\widehat{\mathrm{ETL}}_\alpha(w)$ is

$$\max\ p^{\mathsf{T}}\tilde{w}, \quad \text{subject to} \quad -\frac{1}{L}e \le p \le 0, \quad p^{\mathsf{T}}e = -1, \tag{4.4.6}$$

where $\tilde{w} = Rw$, $L = \alpha T$ and L is assumed to be an integer. Then the optimal solution p^* functions as picking the L smallest elements of $\tilde{w}$. Let $\tilde{w}_{s(l)}$ be the lth smallest element of $\tilde{w}$ counting multiplicities, and define index sets $J^<(\tilde{w},l)$, $J^=(\tilde{w},l)$ and $J^>(\tilde{w},l)$ as follows

$$J^<(\tilde{w},l) = \{i : \tilde{w} < \tilde{w}_{s(l)}\} \tag{4.4.7}$$

$$J^=(\tilde{w},l) = \{i : \tilde{w} = \tilde{w}_{s(l)}\} \tag{4.4.8}$$

$$J^>(\tilde{w},l) = \{i : \tilde{w} > \tilde{w}_{s(l)}\} \tag{4.4.9}$$

Accordingly, $J^{\leq}(\tilde{w}, l) = J^{<}(\tilde{w}, l) \cup J^{=}(\tilde{w}, l)$. Also let $N^{<}(l)$, $N^{=}(l)$ and $N^{>}(l)$ represent the cardinality of corresponding sets respectively. Then the optimal solution set is given by

$$
\begin{aligned}
P(\tilde{w}, L) :=& && (4.4.10)\\
&\{p : p_i = -\frac{1}{L}, \quad \text{for } i \in J^{<}(\tilde{w}, L); && (4.4.11)\\
&\quad p_i = -\tfrac{L-N^{<}(L)}{L}\lambda_i, \sum_{i \in J^{=}(\tilde{w}, L)} \lambda_i = 1, 0 \leq \lambda_i \leq \tfrac{1}{L-N^{<}(L)} \text{ for } i \in J^{=}(\tilde{w}, L); && (4.4.12)\\
&\quad p_i = 0, \quad \text{for } i \in J^{>}(\tilde{w}, L)\} && (4.4.13)\\
=& \{p : p_i = -\frac{1}{L}, \quad \text{for } i \in J^{<}(\tilde{w}, L); && (4.4.14)\\
&\quad p_i = -\frac{1}{L}\xi_i, \sum_{i \in J^{=}(\tilde{w}, L)} \xi_i = (L - N^{<}(L)), 0 \leq \xi_i \leq 1 \text{ for } i \in J^{=}(\tilde{w}, L); && (4.4.15)\\
&\quad p_i = 0, \quad \text{for } i \in J^{>}(\tilde{w}, L)\}\ . && (4.4.16)
\end{aligned}
$$

The subdifferential is given by

$$
\begin{aligned}
\partial\widehat{\mathrm{ETL}}_\alpha(w) &= \{R^\mathsf{T} p^* | p^* \in P(\tilde{w}, l)\} && (4.4.17)\\
&\subset -\frac{1}{L}\sum_{i \in J^{>}(\tilde{w}, L)} R_{i\cdot}^\mathsf{T} - \frac{L - N^{<}(L)}{L}\mathrm{Conv}\left\{R_{i\cdot}^\mathsf{T} | i \in J^{=}(\tilde{w}, J)\right\} && (4.4.18)\\
&=: \partial_r\widehat{\mathrm{ETL}}_\alpha(w)\ . && (4.4.19)
\end{aligned}
$$

Set

$$
\Xi := \left\{ \xi \;\middle|\; \begin{array}{c} \xi_i = 1,\ i \in J^{<}(\tilde{w}, L);\ \xi_i = 0,\ i \in J^{>}(\tilde{w}, L); \\ \sum_{i \in J^{=}(\tilde{w}, L)} \xi_i = (L - N^{<}(L)),\ 0 \leq \xi_i \leq 1,\ i \in J^{=}(\tilde{w}, L) \end{array} \right\}\ .
$$

Using the expression (4.4.17) for the directional derivative $\widehat{\mathrm{ETL}}_\alpha'(w; s)$,

$$
\widehat{\mathrm{ETL}}_\alpha'(w; s) = \max_{z \in \partial\widehat{\mathrm{ETL}}_\beta(w)} \langle z,\ s\rangle\ , \tag{4.4.20}
$$

the condition $\widehat{\text{ETL}}'_\alpha(w;s) \leq 0$ becomes

$$\max_{\xi \in \Xi} -\frac{1}{L}\left[\sum_{j=1}^{T} \xi_j R_{j\cdot} s\right] \leq 0 \ ,$$

or equivalently,

$$\left[\max_{\substack{0 \leq \xi_i \leq 1 \ i \in J^=(\tilde{w},L) \\ \sum_{i \in J^=(\tilde{w},L)} \xi_i = (L - N^<(L))}} \sum_{i \in J^=(\tilde{w},L)} -\xi_j R_{i\cdot} s\right] \leq \sum_{i \in J^<(\tilde{w},L)} R_{j\cdot} s \ .$$

The maximization problem on the left-hand side of this expression is a linear program in the ξ_i's whose dual is the LP

$$\begin{array}{ll} \text{minimize} & (L - N^<(L))\tau + \sum_{i \in J^=(\tilde{w},L)} u_i \\ \text{subject to} & u_i \geq -\tau - R_{i\cdot} s \text{ and } 0 \leq u_i \quad \forall\, i \in J^=(\tilde{w},L) \ . \end{array}$$

Therefore,

$$\left\{ s \mid \widehat{\text{ETL}}'_\alpha(w;s) \leq 0 \right\} = \left\{ s \ \middle| \ \begin{array}{c} \exists\ \tau,\ 0 \leq u_i,\ i \in J^=(\tilde{w},L) \text{ s.t.} \\ u_i \geq -\tau - R_{i\cdot} s,\ i \in J^=(\tilde{w},L) \text{ and} \\ (L - N^<(L))\tau + \sum_{i \in J^=(\tilde{w},L)} u_i \leq \sum_{i \in J^<(\tilde{w},L)} R_{j\cdot} s \end{array} \right\} .$$

Using this fact, the problem $\mathcal{P}_{a2}$ can be rewritten as

$$\underset{s,u,\tau}{\text{maximize}} \quad \widehat{\text{ETG}}'_\beta(w;s) \tag{$\mathcal{P}'_{a2}$}$$

$$\text{subject to} \quad s^{\mathsf{T}} e = 0,\ \|s\|_\infty \leq 1 \tag{4.4.21}$$

$$0 \leq u_i,\ -\tau - R_{i\cdot} s \leq u_i \text{ for } i \in J^=(\tilde{w},L) \tag{4.4.22}$$

$$(L - N^<(L))\tau + \sum_{i \in J^=(\tilde{w},L)} u_i \leq \sum_{i \in J^<(\tilde{w},L)} R_{j\cdot} s \ . \tag{4.4.23}$$

Note that the resulting system of constraints for $\mathcal{P}'_{a2}$ are linear. However, the objective is problematic. An alternative approach is to replace the underlying subdifferentials in $\mathcal{P}_{a2}$ with their approximations given in (4.2.17) and (4.4.19). Using these approximations to the subdifferntials, we define

$$\begin{array}{rcll} \widehat{\text{ETG}}^\dagger_\beta(w;s) & := & \displaystyle\sup_{z \in \partial_r \widehat{\text{ETG}}_\beta(w)} \langle z,\ s\rangle & \text{and} \\ \widehat{\text{ETL}}^\dagger_\beta(w;s) & := & \displaystyle\sup_{z \in \partial_r \widehat{\text{ETL}}_\beta(w)} \langle z,\ s\rangle \ . & \end{array}$$

Then an approximate direction of steepest descent

$$\max_{\substack{\sum \lambda_j = 1 \\ \lambda_j \geq 0,\ j \in J^{=}(\tilde{w},L)}} -\frac{1}{L}\left[\sum_{i \in J^{<}(\tilde{w},L)} R_{i\cdot}s + (L - N^{<}(L)) \sum \lambda_j R_{j\cdot}s\right] \leq 0 \tag{4.4.24}$$

$$-\frac{1}{L}\left[\sum_{i \in J^{<}(\tilde{w},L)} R_{i\cdot}s + (L - N^{<}(L)) \min_{\substack{\sum \lambda_j = 1 \\ \lambda_j \geq 0,\ j \in J^{=}(\tilde{w},L)}} \sum \lambda_j R_{j\cdot}s\right] \leq 0 \tag{4.4.25}$$

In equation (4.4.25), the minimization gives the smallest element of $\{R_{j\cdot}s | j \in J^{=}(\tilde{w}, L)\}$. Multiply $-L$ on both sides and rearrange the terms, we get

$$(L - N^{<}(L)) \min_{j \in J^{=}(\tilde{w},L)} R_{j\cdot}s \geq -\sum_{i \in J^{<}(\tilde{w},L)} R_{i\cdot}s. \tag{4.4.26}$$

This constraint is equivalent to

$$\left\{ (L - N^{<}(L)) R_{j\cdot}s \geq -\sum_{i \in J^{<}(\tilde{w},L)} R_{i\cdot}s \,\middle|\, j \in J^{=}(\tilde{w}, L) \right\}. \tag{4.4.27}$$

For each $j \in J^{=}(\tilde{w}, L)$, define

$$\hat{p}_{(j,w)} = \sum_{i \in J^{<}(\tilde{w},L)} R_{i\cdot}^{\mathsf{T}} + (L - N^{<}(L)) R_{j\cdot}^{\mathsf{T}}, \tag{4.4.28}$$

then the constraint (4.4.27) can be written as

$$\left\{ \hat{p}_{(j,w)}^{\mathsf{T}} s \geq 0 \middle| \, j \in J^{=}(\tilde{w}, L) \right\}. \tag{4.4.29}$$

Using this constraint in the problem $(\mathcal{P}_{a2})$, we get

$$\max_{s \in \mathbb{R}^n} \quad \widehat{\mathrm{ETG}}'_{\beta}(w; s) \tag{4.4.30}$$

$$\text{s.t.} \quad e^{\mathsf{T}} s = 0,\ \|s\|_{\infty} \leq 1 \tag{4.4.31}$$

$$\hat{p}_{(j,w)}^{\mathsf{T}} s \geq 0, \quad j \in J^{=}(\tilde{w}, L) \tag{4.4.32}$$

Define $\hat{P}_w$ to be the matrix including all $\hat{p}_{(j,w)}^{\mathsf{T}}$ for $j \in J^{=}(\tilde{w}, L)$ such that each column conresponds to each $\hat{p}_{j,w}$, then the constraint (4.4.32) is just $\hat{P}_w^T s \geq 0$. Notice that this constraint doesn't affect the reformulation of the objective function $\widehat{\mathrm{ETG}}'_\beta(w;s)$ into the form in (4.3.11). All that needs to be done is to add this constraint in the inner maximization of (4.3.11). So the problem $(\mathcal{P}_{a2})$ is equivalent to the following problem,

$$\max_{j \in I^=(\tilde{w},K)} \max_{\substack{e^\mathsf{T} s=0,\ \|s\|_\infty \leq 1 \\ \hat{P}_w^\mathsf{T} s \geq 0}} \hat{q}_{(j,w)}^\mathsf{T} s, \tag{4.4.33}$$

where $\hat{q}_{(j,w)}$ is the same as in (4.3.11). Observe the inner maximization problem, the constraint $\|s\|_\infty \leq 1$ is the same as a box constraint $-1 \leq s \leq 1$. This makes it a linear programming problem, this is why we impose infinity norm constraint because it enable us to solve it efficiently using existing mature algorithms.

4.5 Backtracking line search

From previous section, we can calculate the steepest ascent direction of $\widehat{\mathrm{ETG}}_\beta(w^k)$, when w^k is on the relative boundary of the feasible set. Denote this steepest ascent direction by Δw^k. This section proposes a backtracking line search method based on w^k and Δw^k to determine how to move to the next iteration such that $\widehat{\mathrm{ETG}}_\beta(w^{k+1})$ is increased by an enough amount.

We first propose a subroutine $\texttt{readBD}(w_c, d)$. It takes two arguments: w_c, a point in the relative interior of the feasible set $\Omega(\sigma_p)$; d, a direction in the subspace spanned by the feasible set. It returns a point w_{bd}, which resides on the relative boundary of the feasible set such that

$$w_{bd} = w_c + kd \tag{4.5.1}$$

for some positive constant k. This is illustrated in Figure 4.2.

At iteration k, let w_c be the vector of weights that gives minimum $\widehat{\mathrm{ETL}}_\alpha$, w^k be the current iterate on the relative boundary of the feasible set $\Omega(\sigma_p)$ (3.4.22), Δw^k be the steepest ascent direction of $\widehat{\mathrm{ETG}}_\beta(w)$ at w^k within the feasible set, and a^k be the corresponding directional derivative. Then the algorithm for the backtracking line search method is stated as follows

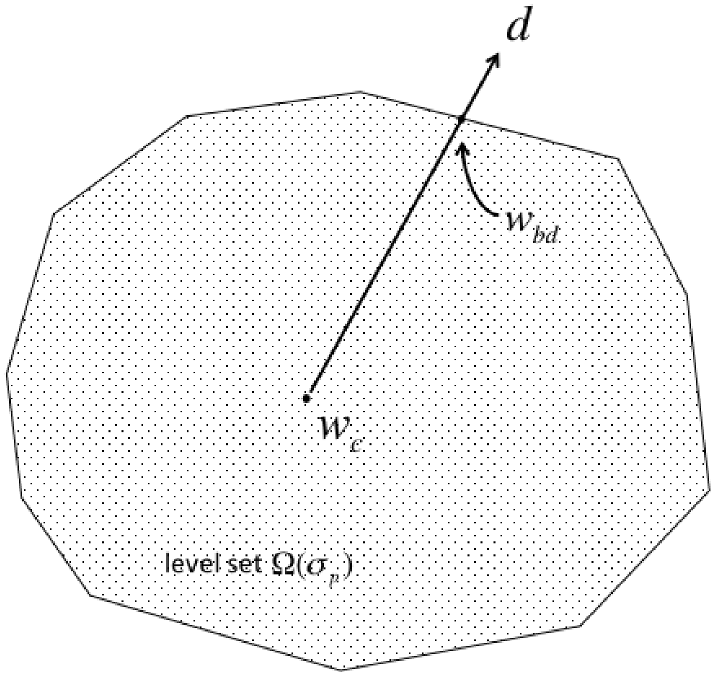

Figure 4.2: Illustration of `reachBD`

Algorithm 1. Backtracking line search method.

given w_c, w^k, Δw^k and a^k.

initialize $t := 1$, $\epsilon > 0$.

if $a^k < \epsilon$, **return** $w^{k+1} = w^k$.

repeat

$w^{temp} = \texttt{reachBD}(w_c, w^k + t\Delta w^k - w_c)$.

$t := t/2$.

until $\left\| w^{temp} - w^k \right\|_2 < \epsilon$ **or** $\widehat{\text{ETG}}_\beta(w^{temp}) > \widehat{\text{ETG}}_\beta(w^k)$.

return $w^{k+1} = w^{temp}$.

Figure 4.3 illustrates one step in `backTrack`. First, it moves from w^k towards Δw^k for length t and reaches $w^k + t\Delta w^k$; next, it calls $\texttt{reachBD}(w_c, w^k + t\Delta w^k - w_c)$ to get the point w^{temp}. If the $\widehat{\text{ETG}}$ value at w^{temp} is greater than the $\widehat{\text{ETG}}$ value at

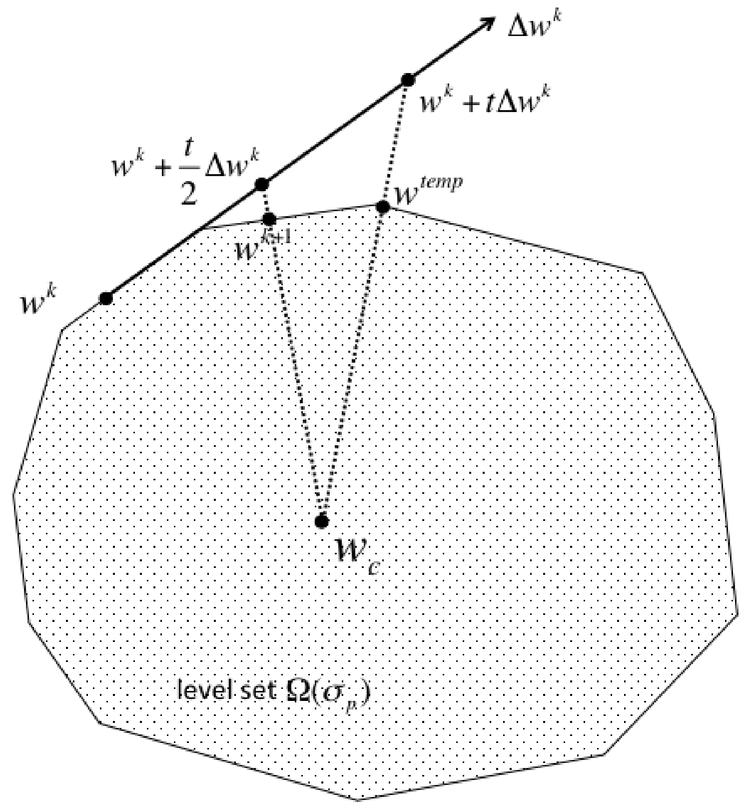

Figure 4.3: Illustration of `backTrack`

w^k, let $w^{k+1} = w^{temp}$ and return it; if not, shrink t's value by a half and repeat the process until a greater $\widehat{\mathrm{ETG}}$ is achieved or t is so small that the distance between w^k and w^{temp} is smaller than the tolerence.

Next two subsections give two algorithms to implement the `reachBD` subroutine: the bisection method and the secant method. The secant method is proved to be super-linear convergent. A comparison of the running time using these two algorithms are discussed in Chapter 5.

4.5.1 Bisection method

For a fixed weight vector w_c and an ascent direction d, define function $f : \mathbb{R}^+ \to \mathbb{R}$ as

$$f(k) = \widehat{\mathrm{ETL}}_\alpha(w_c + kd). \tag{4.5.2}$$

Since $\widehat{\mathrm{ETL}}_\alpha(w)$ is convex in w, as proved in section 4.1, $f(k)$ is also convex. Choose w_c such that it gives the minimum $\widehat{\mathrm{ETL}}_\alpha$, then $f(0) = \widehat{\mathrm{ETL}}_\alpha(w_c) \leq f(k)$ for any $k > 0$.

Claim 2. $f(k)$ is a convex and non-decreasing function.

Proof. 1. Convexity.

For $k_1, k_2 > 0$, $0 < \lambda < 1$,

$$\begin{aligned} f(\lambda k_1 + (1-\lambda)k_2) &= \widehat{\mathrm{ETL}}_\alpha(w_c + (\lambda k_1 + (1-\lambda)k_2)d) && (4.5.3)\\ &= \widehat{\mathrm{ETL}}_\alpha((\lambda(w_c + k_1 d) + (1-\lambda)(w_c + k_2 d)) && (4.5.4)\\ &\leq \lambda\widehat{\mathrm{ETL}}_\alpha((w_c + k_1 d) + (1-\lambda)\widehat{\mathrm{ETL}}_\alpha((w_c + k_2 d) && (4.5.5)\\ &= \lambda f(k_1) + (1-\lambda)f(k_2). && (4.5.6) \end{aligned}$$

We have inequality (4.5.5) because $\widehat{\mathrm{ETL}}_\alpha(w)$ is convex in w.

2. Non-decreasing property.

Suppose $0 < k_1 < k_2$, let $\lambda = \frac{k1}{k2}$, then $0 < \lambda < 1$, we have

$$\begin{aligned} f(k_1) &= f(\lambda k_2 + (1-\lambda)0) && (4.5.7)\\ &\leq \lambda f(k_2) + (1-\lambda)f(0) && (4.5.8)\\ f(k_1) - f(k_2) &= (1-\lambda)(f(0) - f(k_2)) \leq 0 && (4.5.9) \end{aligned}$$

The first inequality is based on convexity of $f(k)$, and the second inequality is based on the property of w_c.

□

Algorithm 2. Bisection method for `reachBD`.

given w_c giving the minimum $\widehat{\mathrm{ETL}}_\alpha$, and a direction d in the subspace spanned by the level set (3.4.22), define $f(k)$ as above,

initialize $l := 0$, $u := u_{max}$ such that $f(u_{max}) > \sigma_p$, tolerence $\epsilon > 0$.

repeat $m := (l + u)/2$. **if** $f(m) < \sigma_p - \epsilon$, $l := m$; **else** $u := m$.

until $\sigma_p - \epsilon < f(m) \leq \sigma$.

An illustration of one step is given in Figure 4.4, where $f(m) < \sigma_p$, so l should be updated as $l = m$.

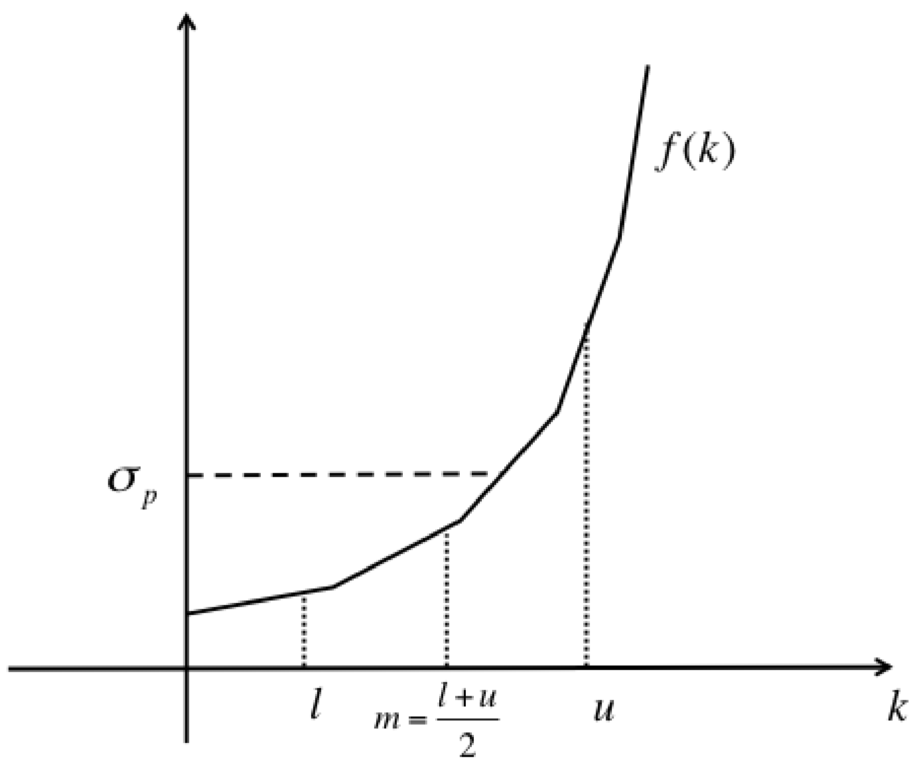

Figure 4.4: Illustration of Bisection method for `reachBD`

4.5.2 Secant method

This section presents an alternative algorithm for `reachBD`. An inexact version of the secant algorithm along with the analysis of rate of convergence can be found in [2]. Still define function $f : \mathbb{R}^+ \to \mathbb{R}$ as

$$f(k) = \widehat{\mathrm{ETL}}_\alpha(w_c + kd), \tag{4.5.10}$$

where w_c is the weight vector giving minimum $\widehat{\mathrm{ETL}}_\alpha$ and d is an ascent direction. For a target $\widehat{\mathrm{ETL}}_\alpha$ value σ_p, define a function `lineRoot(`k_1, k_2`)`. It takes two positive real values, k_1, k_2, as its arguments, computes the intersection of the horizantol line $y = \sigma_p$ and the straight line connecting $(k_1, f(k_1))$ and $(k_2, f(k_2))$, and returns the k coordinate of intersection, namely,

$$\texttt{lineRoot}(k_1, k_2) = k_1 + \frac{\sigma_p - f(k_1)}{f(k_2) - f(k_1)}(k_2 - k_1). \tag{4.5.11}$$

This algorithm starts with two points s_0 and s_1 with $s_0 > s_1$, both are to the right of the root to $f(k) = \sigma_p$. Then according to Claim 2, $f(s_0) > f(s_1) > \sigma_p$. At step k, it takes in s_{k-1} and s_k with $s_k < s_{k-1}$, $f(s_{k-1}) > f(s_k) > \sigma_p$, and returns $s_{k+1} =$`lineRoot`(s_{k-1}, s_k). Since

$$s_{k+1} = s_k + \frac{\sigma_p - f(s_k)}{f(s_{k-1}) - f(s_k)}(s_{k-1} - s_k) < s_k, \tag{4.5.12}$$

the invariant $s_{k+1} < s_k$ is conserved. We can also show $f(s_k) > f(s_{k+1}) > \sigma_p$ is also conserved. Since $f(k)$ is convex, we have

$$f(s_{k+1}) \geq f(s_k) + \gamma_k \frac{\sigma_p - f(s_k)}{f(s_{k-1}) - f(s_k)}(s_{k-1} - s_k), \tag{4.5.13}$$

where γ_k is in the subdifferential of $f(k)$ at s_k, i.e., $\gamma_k \in \partial f(s_k)$. Then

$$\frac{f(s_{k-1}) - f(s_k)}{s_{k-1} - s_k} \geq \gamma_k > 0. \tag{4.5.14}$$

Since $\sigma_p - f(s_k) < 0$, we get

$$\gamma_k \frac{\sigma_p - f(s_k)}{f(s_{k-1}) - f(s_k)}(s_{k-1} - s_k) \geq \sigma_p - f(s_k). \tag{4.5.15}$$

So replace the second term in (4.5.13), we get

$$f(s_{k+1}) \geq \sigma_p. \tag{4.5.16}$$

Theorem 4.5.1. *Suppose that f is defined as in (4.5.10), and the starting points s_0 and s_1 satisfy $f(s_0) > f(s_1) > \sigma_p$ with $s_0 > s_1$. If $f(k) = \sigma_p$ has a root s^*, and the subdifferential of $f(k)$ at s^* is a closed interval $[\alpha, \beta]$ with $\beta > 0$, then*

1. *the iterates generated by the secant method converge to the root of $f(k) = \sigma_p$;*

2. *the rate of convergence of $\{s_k\}$ is superlinear.*

Proof. 1. We first show that s_k is bounded below by s^*. As proved in (4.5.16), for any positive integer k,

$$f(s_k) \geq \sigma_p = f(s^*). \tag{4.5.17}$$

Since $f(k)$ is non-decreasing, we must have $s_k \geq s^*$. Then according to sub-differential inequality, we have for any $\gamma_k \in \partial f(s_k)$ and $\beta \in \partial f(s^*)$,

$$f(s_k) \geq f(s_{k-1}) + \gamma_{k-1}(s_k - s_{k-1}) \tag{4.5.18}$$

$$f(s_k) \geq f(s^*) + \beta(s_k - s^*) \tag{4.5.19}$$

Since $s_{k-1} > s_k > s^*$ and $f(s_{k-1}) > f(s_k) > f(s^*)$, we can get

$$\frac{s_{k-1} - s_k}{f(s_{k-1}) - f(s_k)} \geq \frac{1}{\gamma_{k-1}} > 0 \tag{4.5.20}$$

$$f(s_k) - f(s^*) \geq \beta(s_k - s^*) \geq 0 \tag{4.5.21}$$

Combine these two inequalities, we get

$$\frac{f(s_k) - \sigma_p}{f(s_{k-1}) - f(s_k)}(s_{k-1} - s_k) \geq \frac{\beta}{\gamma_{k-1}}(s_k - s^*) \tag{4.5.22}$$

Then, from equation (4.5.12), we have

$$s_{k+1} - s^* = s_k - s^* + \frac{\sigma_p - f(s_k)}{f(s_{k-1}) - f(s_k)}(s_{k-1} - s_k) \tag{4.5.23}$$

$$\leq (1 - \frac{\beta}{\gamma_{k-1}})(s_k - s^*) \tag{4.5.24}$$

Recall that $\beta \in \partial f(s^*)$ and $f(k)$ is convex, so $\gamma_0 > \gamma_{k-1} \geq \beta > 0$ for any $\gamma_{k-1} \in \partial f(s_{k-1})$ when $s_{k-1} > s^*$, $k > 1$. Then $1 - \frac{\beta}{\gamma_{k-1}} < 1 - \frac{\beta}{\gamma_0} < 1$. Hence,

$$\frac{|s_{k+1} - s^*|}{|s_k - s^*|} \leq 1 - \frac{\beta}{\gamma_0} < 1. \tag{4.5.25}$$

This proves that s_k converges to s^* linearly.

2. Given that s_k converges to s^*, we have $\gamma_k \to \beta$ as $k \to \infty$. So, from inequality (4.5.24),

$$\lim_{k\to\infty} \frac{|s_{k+1} - s^*|}{|s_k - s^*|} = 0. \tag{4.5.26}$$

This proves that s_k converges to s^* superlinearly.

$\square$

Algorithm 3. Secant method for `reachBD`.

given w_c giving the minimum $\widehat{\mathrm{ETL}}_\alpha$, and an ascent direction d in the subspace spanned by the level set (3.4.22), define $f(k)$ as above,

initialize $s_1 := 1$, tolerence $\epsilon > 0$.

repeat `(Phase one)`

  if $f(s_1) > \sigma_p$, **break**;

  else $s_1 := 2s_1$;

end repeat

$s_0 := 2s_1$;

repeat `(Phase two)`

  $\hat{s} :=$ `lineRoot`(s_0, s_1);

  if $|s_0 - s_1| < \epsilon$

    return s_0;

  if $\sigma_p - \epsilon < f(\hat{s}) < \sigma_p$

    return $\hat{s}$;

  $s_0 := s_1$;

$s_1 := \hat{s}$;

end repeat

Phase one starts with an initial step size $s_1 = 1$, and double it until s_1 is to the right of the root, i.e., $f(s_1) > \sigma_p$, then it uses $2s_1$ as the value of s_0 such that $s_1 > s_0$ and $f(s_1) > f(s_0) > \sigma_p$. Figure 4.5 shows an illustration of phase two. It starts with two points s_0 and s_1, both are to the right of the root, the method described above to find $\hat{s}$ guarantees it is in between the root and s_1. Let $s_0 = s_1$ and $s_1 = \hat{s}$ and repeat this process, $\{s_k\}$ are guranteed to converge to the root superlinearly.

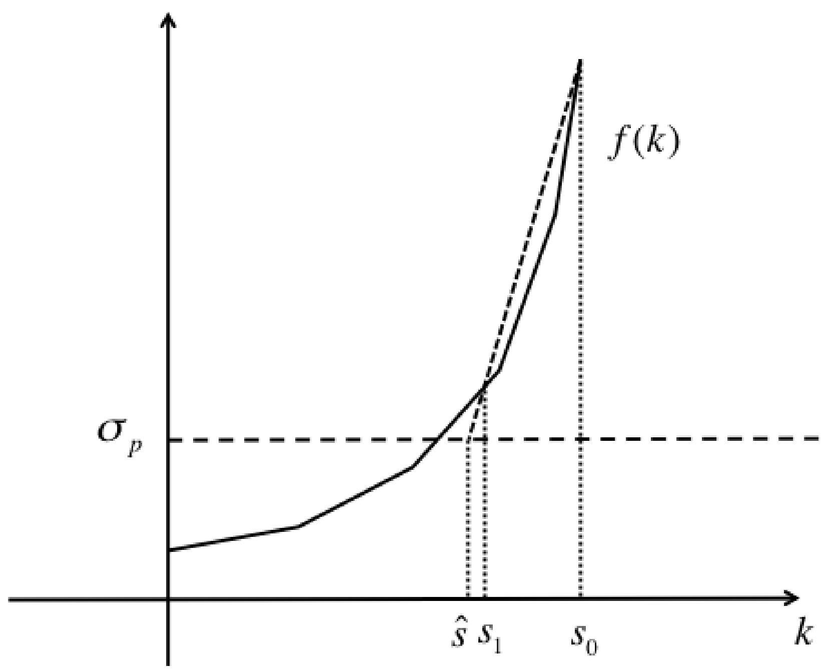

Figure 4.5: Illustration of Secant method for `reachBD`

4.6 The algorithm scheme

1. Initialize the starting point w^0 in the feasible set.

Solve the following problem

$$\begin{aligned} \underset{w}{\text{minimize}} \quad & \widehat{\mathrm{ETL}}_\alpha(w) \\ \text{subject to} \quad & w^\mathsf{T} e = 1 \end{aligned}$$

Denote its optimal solution by w_{mETL}. Let $w^0 = w_{mETL}$, then $\widehat{\mathrm{ETL}}_\alpha(w^0)$ gives the minimum $\widehat{\mathrm{ETL}}$ for this portfolio. w^0 is guaranteed to be in the feasible set of the ETG $-$ ETL problem, or the problem is infeasible when the target ETL σ is less than $\widehat{\mathrm{ETL}}_\alpha(w^0)$.

2. Construct a set of initial directions starting from w^0 that span the subspace containing the feasible set. We choose the directions such that they form an orthogonal coordinate system in the subspace, with w^0 as the origin. This can be achieved by QR decomposition. First solve the following problem

$$\begin{aligned} \underset{w}{\text{minimize}} \quad & \widehat{\mathrm{ETG}}_\beta(w) \\ \text{subject to} \quad & w^\mathsf{T} e = 1 \end{aligned}$$

and denote its optimal solution by w_{mETG}. Then define $d_1 = w_{mETG} - w_{mETL}$ and define matrix X , which contains e and d_1 as its two columns, as follows,

$$X = (e, d_1)_{n \times 2}. \tag{4.6.1}$$

Here e is the vector of 1's and n is the number of assets. Notice that

$$e^\mathsf{T} d_1 = e^\mathsf{T} w_{mETG} - e^\mathsf{T} w_{mETL} = 1 - 1 = 0, \tag{4.6.2}$$

so e and d_1 are orthogonal to each other. Take a QR decomposition of X and complete it by adding the set of arbitrary orthogonal vectors, $d_2, \ldots, d_{n-1}$, from the QR decomposition, we get

$$X_{full} = (e, d_1, \ldots, d_{n-1})_{n \times n}. \tag{4.6.3}$$

Then $\{d_1, \ldots, d_{n-1}\}$ forms an orthogonal base for the subspace containing the feasible set. In the R code, the construction of X_{full} is by function `qr` and `qr.X` in the base package. The QR decomposition is computed by LINPACK or LAPACK [35, 9, 3].

3. Along each direction in the set $\{d_1, \ldots, d_{n-1}\}$, move one step forward to w^1 but still within the relative interior of the feasible set. Then follow along the relaxed steepest ascent direction of $\widehat{\text{ETG}}_\beta$ at w^1 to reach relative boundary of the feasible set using `reachBD`. Starting from the point on the relative boundary, take an iterative search along the relaxed steepest ascent direction using `backTrack` at each step until a local maximum is reached.

   Suppose we are at step k. Let $\tilde{w}^k = Rw^k$.

   (a) When $k = 1$, $w^k \in \mathbf{Int}\{\mathbf{w} : \widehat{\text{ETL}}_\alpha(\mathbf{w}) \leq \sigma\}$. For each $j \in I^=(\tilde{w}^k, \beta T)$, construct $e_{(j,w^k)}$ and compute $\left\|P\hat{q}_{(j,w^k)}\right\|_2$ where $P = I - \frac{ee^\mathsf{T}}{n}$ and $\hat{q}_{(j,w^k)} = R^\mathsf{T} e_{(j,w^k)}$. Find j^* that gives the maximum value of $\left\|P\hat{q}_{(j,w^k)}\right\|_2$, then

   $$\Delta w^k = \frac{P\hat{q}_{(j^*,w^k)}}{\left\|P\hat{q}_{(j,w^k)}\right\|_2}$$

   gives the relaxed steepest ascent direction.

   (b) When $k > 1$, $w^k \in \mathbf{bd}\{\mathbf{w} : \widehat{\text{ETL}}_\alpha(\mathbf{w}) \leq \sigma\}$. For each $j \in I^=(\tilde{w}^k, \beta T)$, construct $\hat{q}_{(j,w^k)}$ as in part (3a), also construct $\hat{P}_{w^k}$ as we stated in previous section, then solve the following linear programming problem

   $$\begin{aligned} \max_{s \in \mathbb{R}^n} \quad & \hat{q}_{(j,w^k)}^\mathsf{T} s \\ \text{s.t.} \quad & e^\mathsf{T} s = 0,\ \|s\|_\infty \leq 1 \\ & \hat{P}_{w^k}^\mathsf{T} s \geq 0 \end{aligned}$$

   Find j^* that gives the maximum optimal value $\hat{q}_{(j^*,w^k)}^\mathsf{T} s$, and the corresponding optimal solution s^* give the relaxed steepest ascent direction Δw^k.

(c) Move to w^{k+1} by `backTrack`$(w^k, \Delta w^k)$ introduced in section 4.5.

Since $\widehat{\mathrm{ETG}}_\beta(w)$ is convex, its directional derivative is always greater than or equal to zero, and is strictly positive except at the minimum, thus

$$\begin{aligned}\widehat{\mathrm{ETG}}_\beta(w^k + t\Delta w^k) &\geq \widehat{\mathrm{ETG}}_\beta(w^k) + t\widehat{\mathrm{ETG}}'_\beta(w^l; \Delta w^k)\\ &\geq \widehat{\mathrm{ETG}}_\beta(w^k)\end{aligned}$$

And this inequality is strict when w^k is not at the minimum. This guarantees $\widehat{\mathrm{ETG}}_\beta(w)$'s value is always increasing along Δw.

(d) Repeat step 3 until it reaches a local maximum, i.e., $\widehat{\mathrm{ETG}}'_\beta(w^k, \Delta w) < \epsilon$, or when the distance between two iterates w^k and w^{k+1} is small enough, i.e., $\left\|w^{k+1} - w^k\right\|_2$, or when ETG cannot increase in an amount larger than the torlerance, i.e., $|\widehat{\mathrm{ETG}}_\beta(w^k + t\Delta w^k) - \widehat{\mathrm{ETG}}_\beta(w^k)| < \epsilon$.

4. Choose the maximum of the $2n - 2$ ETG values, and return the corresponding local optimal solution as the optimal weight vector of the algorithm.

Chapter 5
NUMERICAL EXPERIMENTS

In this chapter, we report the performance of the proposed ETG-ETL optimization algorithm. According to the nature of the problem, there can be infinite number of local optimal solutions, all of which are on the relative boundary of the feasible set. This algorithm tries to search along the directions that span the subspace defined by the feasible set, such that it can cover the relative boundary as much as possible. Although it is not guaranteed to converge to a global optimal solution, our numerical implementation shows expected result in an out of sample backtesting experiment, and enables potential merit of usage in practice.

This chapter is organized as follows: Section 5.1 discusses a simple example using three assets. This choice of small number of assets in the portfolio enables us to visualize in the three dimensional space and facilitates us to present clearly the behaviour of the algorithm; Section 5.2 gives an analysis on the local convergence rate; Section 5.3 applys the algorithm in a realistic setting: a portfolio consisting of the 30 member stocks of Russell 2000. We optimize the portfolio using classical mean-variance optimization, mean-ETL optimization and the ETG-ETL optimization on daily and monthly bases using different values for parameters α and β, and compare the cumulative returns of the portfolios with initial value $1.

5.1 An numerical example using three assets

This example considers a portfolio of three stocks: "AMZN", "ORCL" and "LLTC". It uses daily adjusted returns from January 25th, 2008 to December 31st 2010. We choose the dates such that there are exactly $T = 740$ periods in total. The data is downloaded from Yahoo Finance using function `get.hist.quote()` in R package `tseries`.

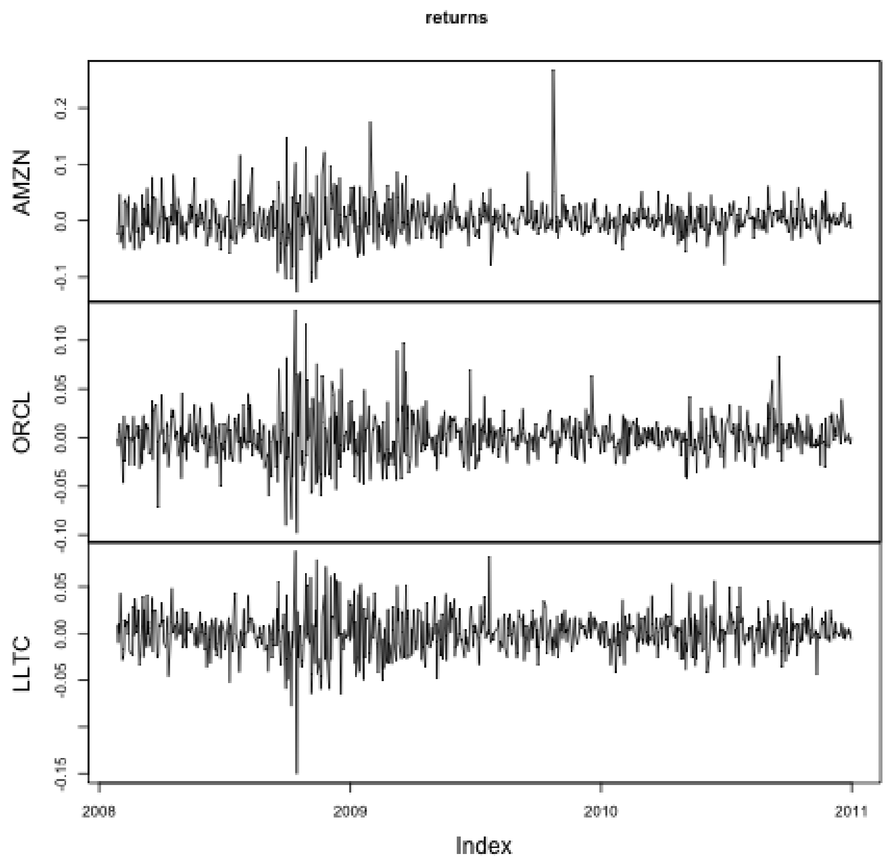

Figure 5.1: Returns time series for AMZN, ORCL and LLTC

Figure 5.1 shows the plot of the returns time series and Figure 5.2 shows the qq-plot and kernel density estimation respectively. From both types of plots, we can see that none of the three stocks shows normality. All of them have fat tails.

We show the searching paths of the algorithm in two cases:

1. $\alpha = 0.05$ and $\beta = 0.05$.

2. $\alpha = 0.05$ and $\beta = 0.5$.

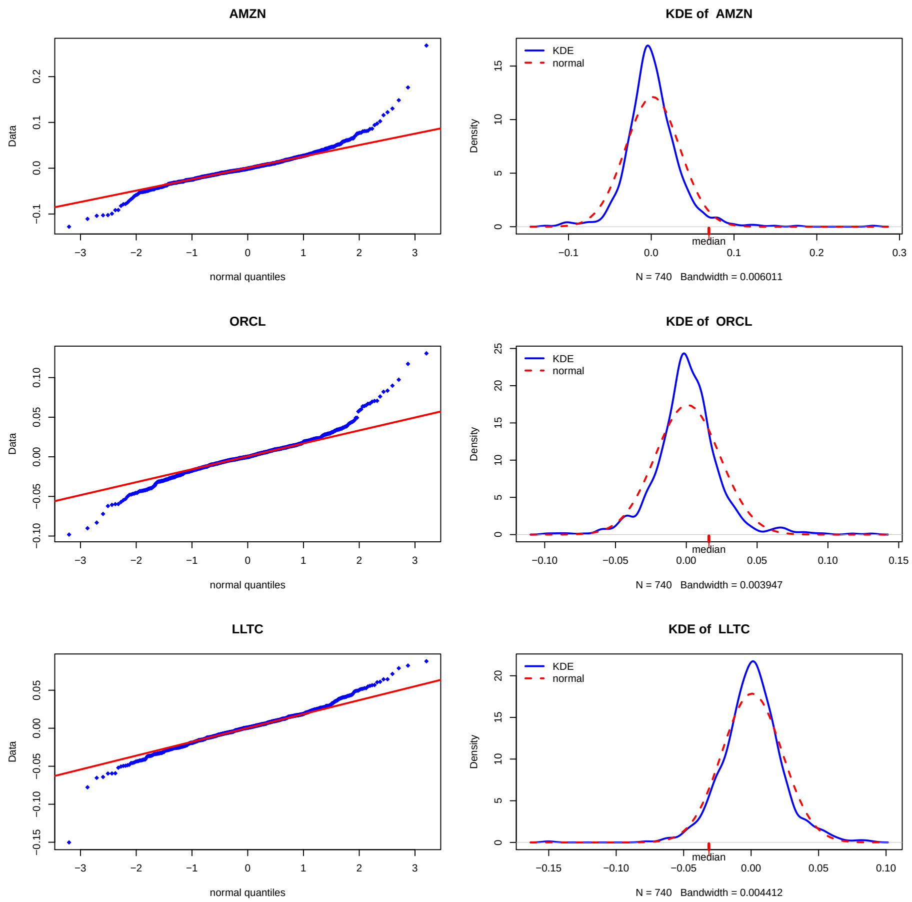

Figure 5.2: QQ-plots and kenel density estimations for AMZN, ORCL and LLTC

In both cases, α and β are chosen such that αT and βT are integers. Figure 5.3 and figure 5.4 show the searching paths corresponding to case 1 and case 2 respectively.

In both three dimentional plots, the three axis are weights w_1, w_2 and w_3 respectively. The shaded region represents the feasible set, i.e. the sublevel set $\widehat{\text{ETL}} \leq \sigma$, and the black curve is the a contour line of the $\widehat{\text{ETG}}$. These two sets are generated by computing $\widehat{\text{ETL}}$ and $\widehat{\text{ETG}}$'s values on a gird of values for $w = (w_1, w_2, w_3)^\intercal$ in a bruteforce way. The point labeled by $w.mETL$ is the weight vector that gives the minimum ETL value, and the point labeled by $w.mETG$ is the weight vector that gives the minimum ETG value. Observe that the sublevel set $\widehat{\text{ETL}} \leq \sigma$ is a polyhe-

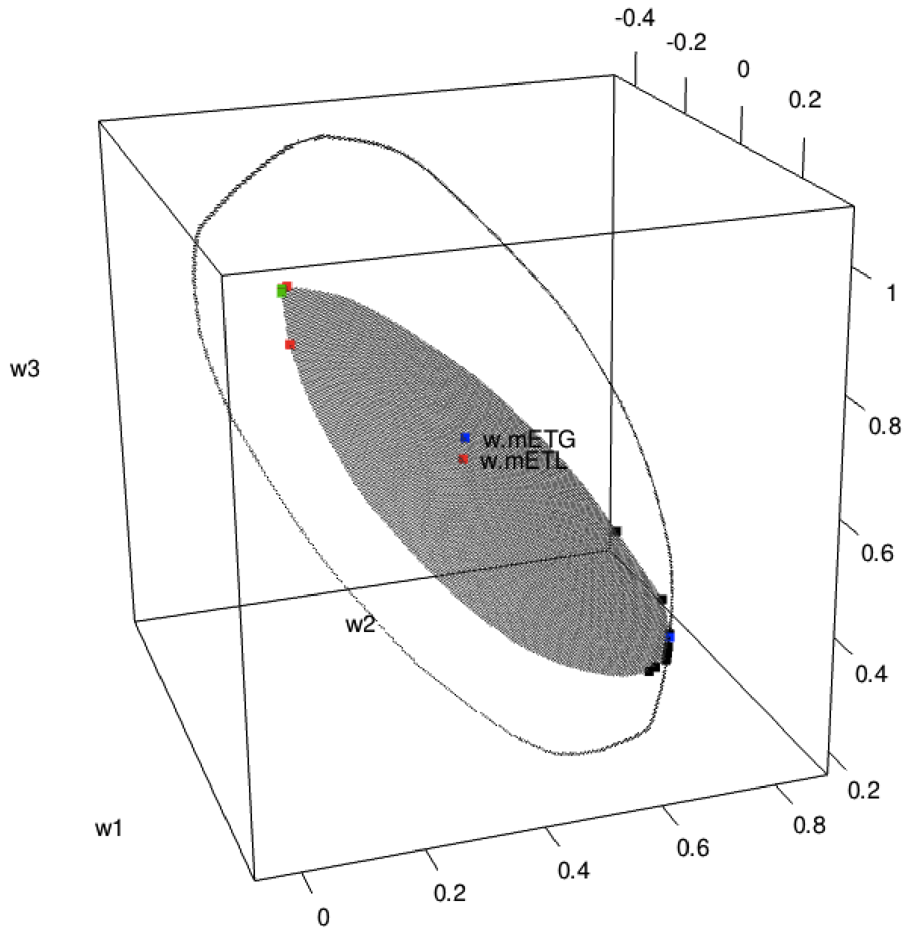

Figure 5.3: Searching paths for case 1.

dron and both the sublevel set and the $\widehat{\mathrm{ETG}}$ contour line are in the plane defined by $w^{\mathsf{T}}e = 1$, as we expect them to be. It is clear that the optimal solution is the point where these two sets intersects.

The algorithm searches along four directions, two of which are along the line that goes through $w.mETL$ and $w.mETG$, the other two directions along the direction that is perpendicular to it in the subspace defined by $w^{\mathsf{T}}e = 1$. The four trajectories are marked by black, red, green and blue.

In Figure 5.3, the black and blue trajectories lead to the global maximum, but the red and green trajectories are trapped at a local maximum. Table 5.1 shows the numerical values for the optimal w, ETG value, maximum ascent value at the optimal w and last stepsize for each direction, where the tolerance in Algorithm 1 is set to be $\epsilon = 8.161993^{-15}$.

Table 5.1: Optimal w, ETG, maximum ascent value at the optimal w and last stepsize for case 1.

	w1	w2	w3	ETG	max.ascent	last stepsize
black	0.017994	0.770014	0.211992	0.052424	4.172660E-12	1.455190E-11
red	-0.040385	0.117905	0.922480	0.050104	0.000000E+00	4.768370E-07
green	-0.040385	0.117905	0.922480	0.050104	0.000000E+00	2.384190E-07
blue	-0.020468	0.793748	0.226720	0.052430	6.057380E-13	7.105430E-15

In Figure 5.4, all the trajectories lead to the global maximum except the black trajectory, which is trapped at a local maximum in the opposite direction. Table 5.2 shows the numerical values for the optimal w, ETG value, maximum ascent value at the optimal w and last stepsize for each direction, where the tolerance in Algorithm 1 is set to be $\epsilon = 8.161993^{-15}$.

Table 5.2: Optimal w, ETG, maximum ascent value at the optimal w and last stepsize for case 2.

	w1	w2	w3	ETG	max.ascent	last stepsize
black	-0.069978	0.805769	0.264209	0.015832	7.725380E-12	7.105430E-15
red	-0.040385	0.117905	0.922480	0.016559	0.000000E+00	4.768370E-07
green	-0.040385	0.117905	0.922480	0.016559	0.000000E+00	4.768370E-07
blue	-0.040385	0.117905	0.922480	0.016559	0.000000E+00	4.768370E-07

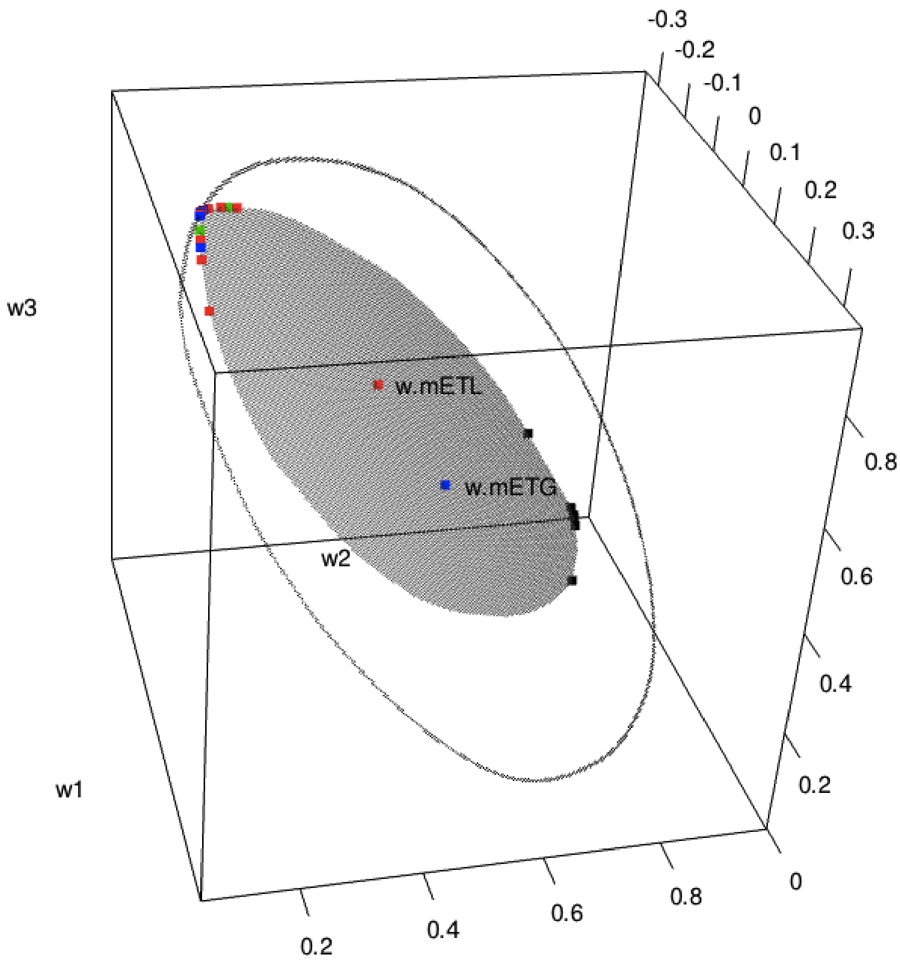

Figure 5.4: Searching paths for case 2.

Figure 5.5 gives a comparison of the optimal weights for case 1 and case 2. Observe that as β increases from 0.05 to 0.5, more weight are moved from asset 2, "ORCL", to asset 3, "LLTC". This is confirmed by checking the kernel density estimation plot in Figure 5.2. Observe that the average value for returns above the median is much larger for "LLTC" than "ORCL".

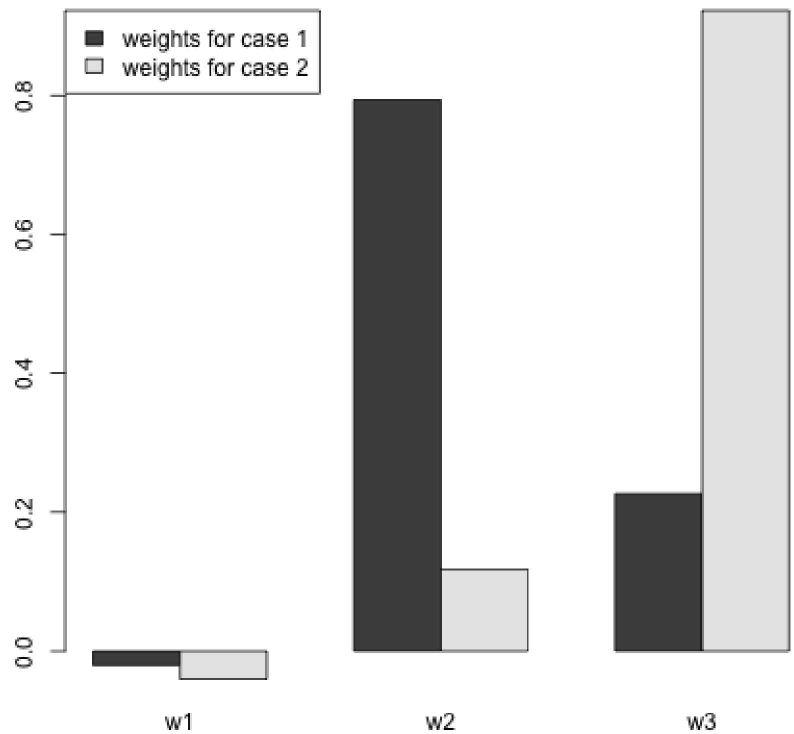

Figure 5.5: Weights for case 1 and case 2.

5.2 Convergence analysis

This section checks the relationship between the numerical maximum ascent values at all the local optimal solutions and the tolerance in the stopping criteria. Theoretically, the maximum ascent is equal to 0 at a local maximum. Numerically, we consider a number 0 when it is smaller than a torlerance, usually a very small number. As indicated in section 4.6, we implement the algorithm such that it stops when the numerical value for the maximum ascent, i.e., the directional derivative $\widehat{\mathrm{ETG}}'_{\beta}(w; \Delta w)$, is less than a torlerance ϵ, or when the distance between two iterates is less than ϵ,

$\left\|w^{k+1} - w^k\right\|_2 < \epsilon$ or $|\widehat{\mathrm{ETG}}(w^{k+1}) - \widehat{\mathrm{ETG}}(w^k)| < \epsilon$, no matter the maximum ascent is close to 0 or not. However, as the torlerance ϵ decreases, we are able to reach to the local maximums whose maxium ascent values are close to 0, but it also takes longer CPU time to achieve a better precision. We show these relationships by analyzing a portfolio of 5 assets using a sequence of torlerance values. We use adjusted daily returns from January 4th, 2006 to December 30th, 2011, for "AA", "AXP", "BA", "BAC" and "CAT" and optimize the portfolio using $\alpha = 0.05$, $\beta = 0.5$. The result is illustrated in Figure 5.6. For each torlerance value, we plot the boxplot of all the

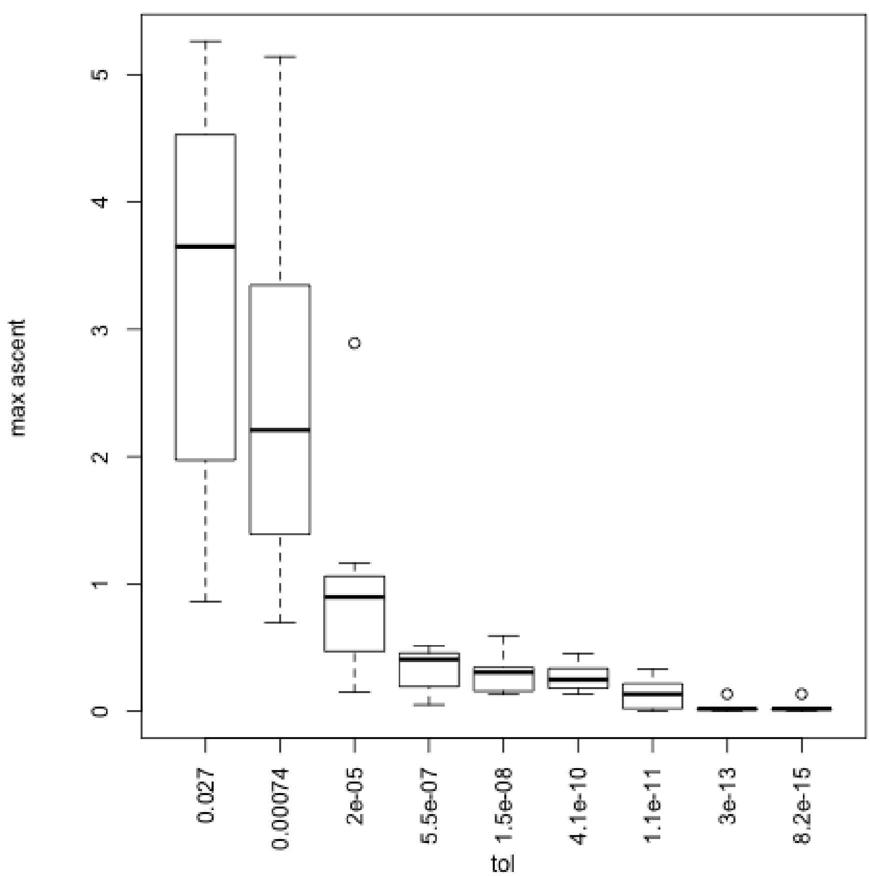

Figure 5.6: Boxplot of the max ascent values v.s. tolerance.

maximum ascent values at all the 8 local optimal solutions. With other conditions remain, as the tolerance decreases from 0.027 to 8.2×10^{-15}, the maximum ascent values not only move towards 0 but also become more gathered. Figure 5.7 shows the

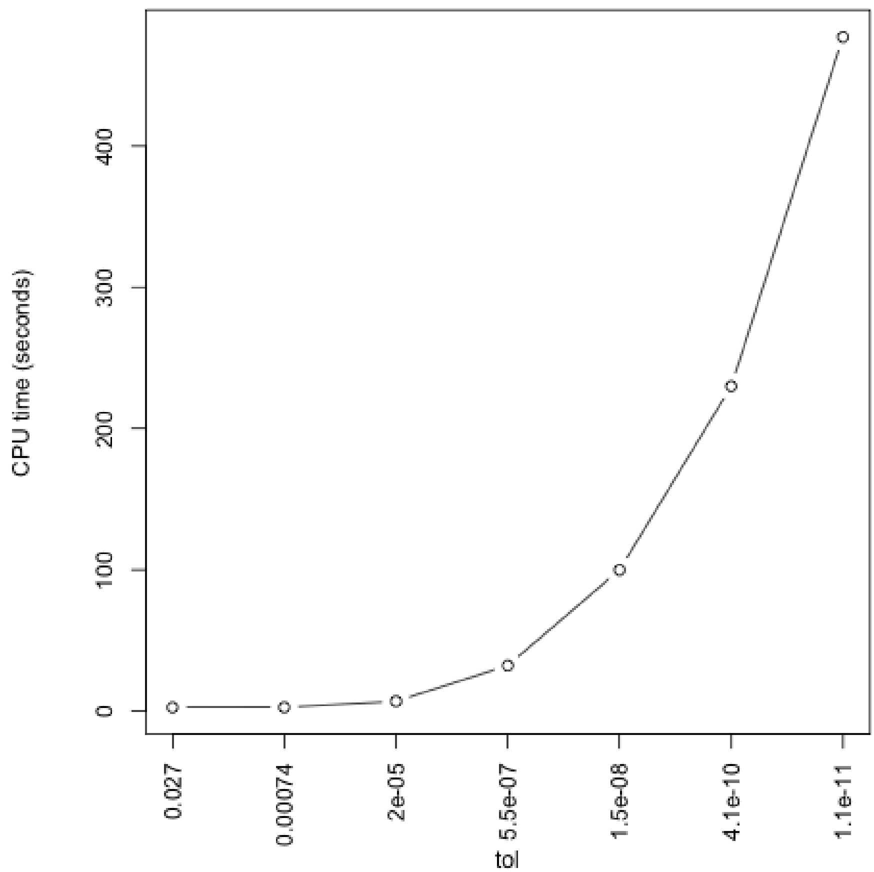

Figure 5.7: CPU time v.s. tolerance.

relationship between the CPU time and the torlerance levels. Starting from torlerance value smaller than 5.5×10^{-7}, the CPU time increase rapidly. Most of the CPU time is spent to gain the precision smaller than 5.5×10^{-7}. Practically, we consider a tolerance at 10^{-6} to be reasonable. On one hand, This means a minimum \$1 change for a capital of \$1 million, which is less than one share for a usual stock price. On the other hand, this enable a fast computation of the portfolio weights. The algorithm was carried out using R code and `lpSolve` package for the linear programming subroutines on a 1.7 GHz Intel Core i5 machine. It took 300 seconds on optimizing a portfolio of 100 stocks, and 1929 seconds for a portfolio of 470 stocks, which were selected from the member stocks of S&P 500 index, excluding the other 30 member stocks because there is not enough historical data.

5.3 A more realistic example

This section applies the algorithm on a portfolio of 30 stocks selected from the member stocks of Russell 2000, which measures the performance of the small-cap segment of the U.S. equity universe. It represents approximately 10% of the total market capitalization of the Russell 3000. These 30 member stocks are: BGFV, STRL, ALNY, APAGF, MHR, NATL, SYBT, COCO, FMD, ACO, AZZ, ARUN, EBS, TLEO, FCN, THRX, BKE, SONO, ARC, LGND, PVA, EHTH, ASEI, EXR, TDY, PNY, PLFE, LCRY, KOP, SONE. It uses the daily adjusted closed prices from January 2nd, 2008 to December 30th, 2011. There are totally $T = 1008$ periods. The data is downloaded from Yahoo Finance using function `get.hist.quote()` in R package `tseries`.

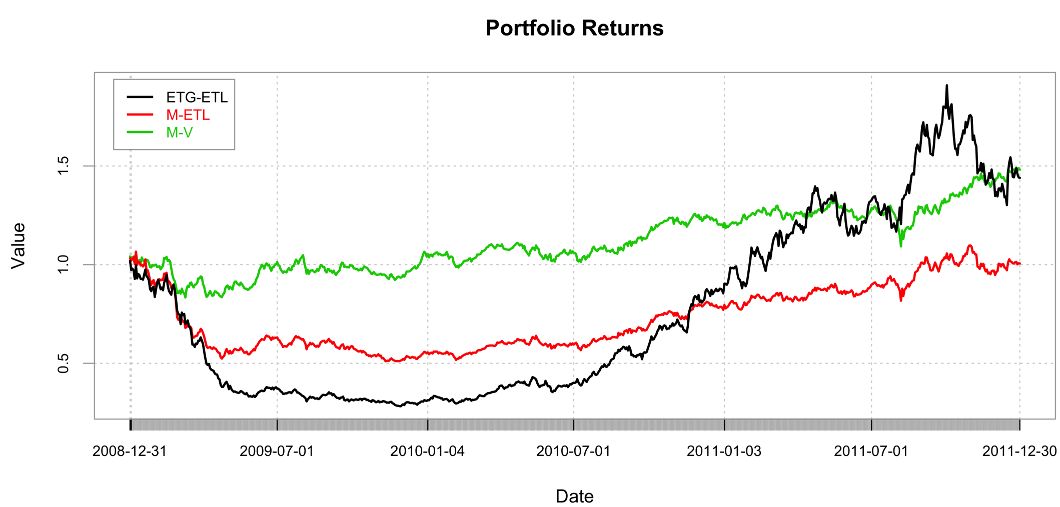

Figure 5.8: \$1 portfolio cumulative return. Daily rebalance. $\alpha = 0.05$, $\beta = 0.95$, $\sigma_p = \min \mathrm{ETL} + 0.01$.

Figure 5.8 shows the dollar value of a portfolio with initial value \$1 and rebalanced daily using one year training data. It optimizes the portfolio using three methods: the classical mean-variance optimization, the mean-ETL optimization and the $\mathrm{ETG}-\mathrm{ETL}$ optimization. To make the $\mathrm{ETG}-\mathrm{ETL}$ optimization problem feasible, we choose the target ETL to be $\sigma_p = \min \mathrm{ETL} + 0.01$. The performance of the portfolio using the $\mathrm{ETG}-\mathrm{ETL}$ optimization method is the worst. This is because that the ETL value is high during the 2008-2009 finanial crisis period, thus making the target ETL too large to be a reasonable risk preference. To remedy this, we fix the target ETL to be 0.02 in the $\mathrm{ETG}-\mathrm{ETL}$ optimization. When the problem is not feasible, namely, $\sigma_p < \min \mathrm{ETL}$, as in the financial crisis period, we use the mean-ETL optimization method instead. And the performance is shown in Figure 5.9. The trajectories of the mean-ETL method and the modified $\mathrm{ETG}-\mathrm{ETL}$ method coincide with each other until January 2010. After the crisis, the value of the portfolio using $\mathrm{ETG}-\mathrm{ETL}$ method starts to climb up and cross over the value of the portfolio using the mean-variance method. This indicates the potential merit of the $\mathrm{ETG}-\mathrm{ETL}$ method under

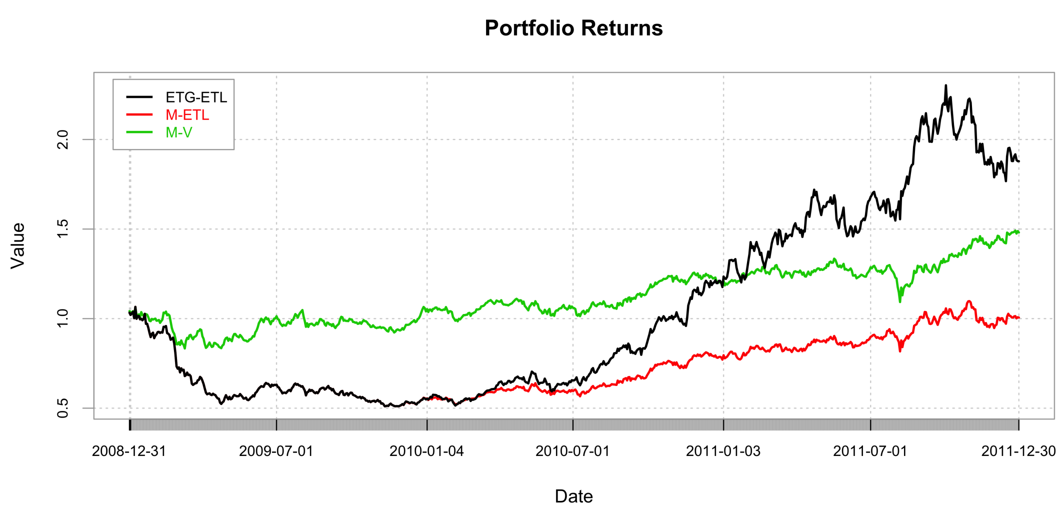

Figure 5.9: \$1 portfolio cumulative return with daily rebalance. $\alpha = 0.05$, $\beta = 0.95$, $\sigma_p = 0.02$.

a normal market condition, but not a crisis period, when the tail loss is more possible to happen. Figure 5.10 compares the three methods during the period after the crisis, from January 4th, 2010 to December 30th, 2011. And the $\mathrm{ETG}-\mathrm{ETL}$ methods outperforms the other two methods in a recognizable amount. This confirms our expectation for the use of the $\mathrm{ETG}-\mathrm{ETL}$ method under a normal market condition. Since daily rebalance would inccur large transaction cost and thus not possible in

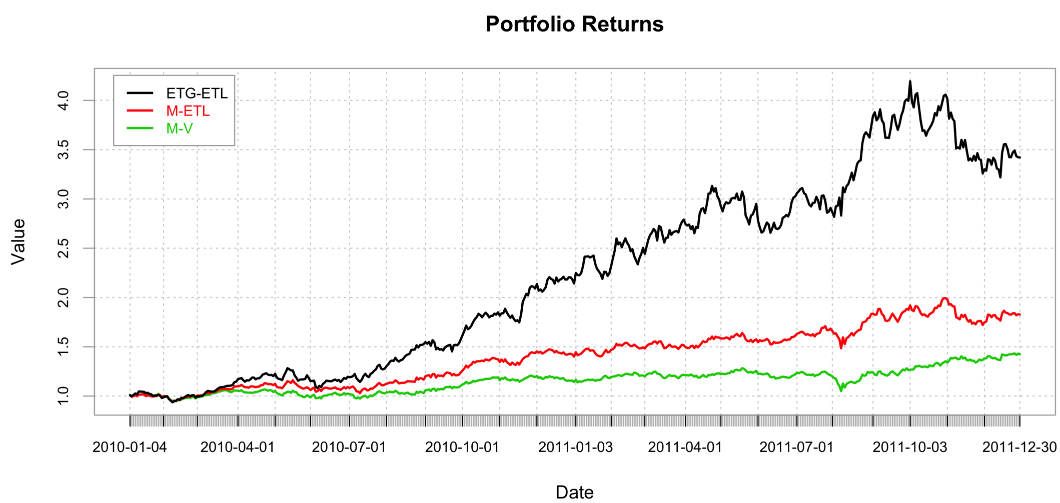

Figure 5.10: \$1 portfolio cumulative return with daily rebalance after the crisis period. $\alpha = 0.05$, $\beta = 0.95$, $\sigma_p = 0.02$.

real-world portfolio management, a less rebalance frequency is more desirable. We did the backtest using monthly rebalance with $\alpha = 0.1$ and $\beta = 0.9$, shown in Figure 5.11 and Figure 5.12. The outperformance of the portfolio using $\mathrm{ETG}-\mathrm{ETL}$ method is still strong on a monthly rebalance base. This enables our algorithm useful in a practical setting.

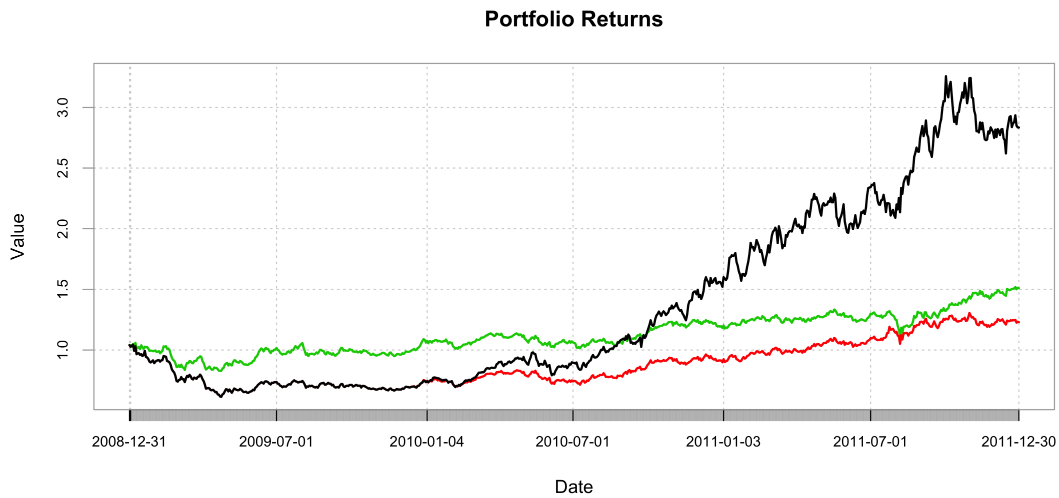

Figure 5.11: \$1 portfolio cumulative return with monthly rebalance. $\alpha = 0.1$, $\beta = 0.9$, $\sigma_p = 0.02$.

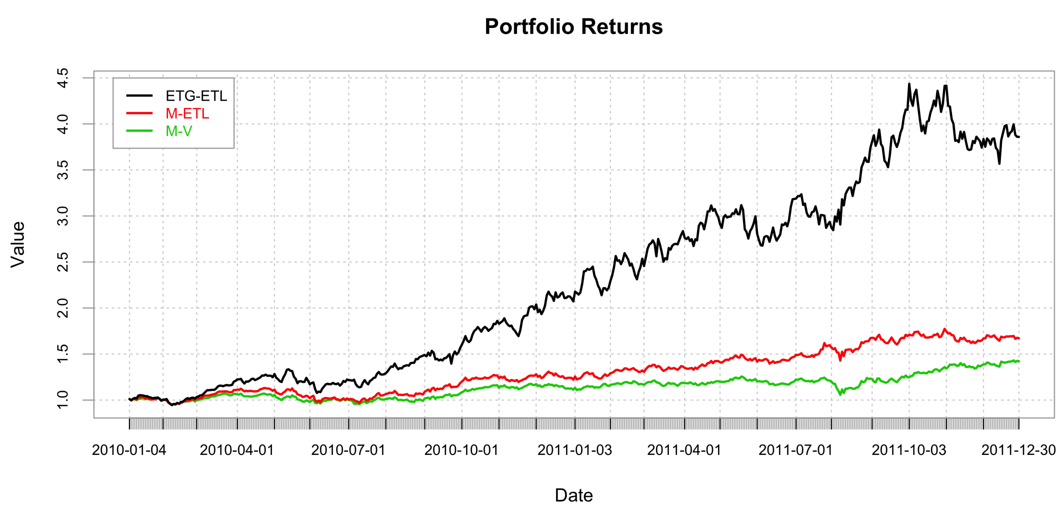

Figure 5.12: \$1 portfolio cumulative return with monthly rebalance after the crisis period. $\alpha = 0.05$, $\beta = 0.95$, $\sigma_p = 0.02$.

5.4 Future Research

In conclusion, our ETG − ETL optimization algorithm appears to have merits of exploring the upper gain potential whil constraining the average tail loss. It is suitable for a normal market condition rather than financial crisis period when downside returns are clustered. An extension to this algorithm is to consider extra constraints such as long only constraint:

$$w \geq 0, \tag{5.4.1}$$

and box constraint:

$$w_{low} < w < w_{high}, \tag{5.4.2}$$

or constraint on subset of w, for example,

$$b_{low} < c^{\mathsf{T}} w < b_{high}, \tag{5.4.3}$$

where c is a column vector with elements to be either 1's or 0's. These are all linear constraints, so the resulting feasible set for the $\mathrm{ETG} - \mathrm{ETL}$ optimization problem is still convex. Then it is worth investigating the expressions for the maximum ascent of $\widehat{\mathrm{ETG}}_{\beta}(w)$ when w is on the relative boundary of the feasible set.

BIBLIOGRAPHY

[1] S. Rachev A. Biglova, S. Ortobelli and S. Stoyanov. Different approaches to risk estimation in portfolio theory. *The Journal of Portfolio Management*, pages 103–112, 2004.

[2] Michel P. Friedlander Aleksandr Y. Aravkin, James V. Burke. Variational properties of value functions and applications to inversion. *Working paper*, June 2012.

[3] E. Anderson. *Lapack Users' Guide*. Software, Environments, Tools. Society for Industrial and Applied Mathematics, 1999.

[4] Vijay S. Bawa. Optimal rules for ordering uncertain prospects. *Journal of Financial Economics*, 2(1):95–121, 1975.

[5] J. Brodie, I. Daubechies, C. De Mol, D. Giannone, and I. Loris. Sparse and stable Markowitz portfolios. Technical report, European Central Bank, Working Paper No. 936, 2008.

[6] Giuseppe Carlo Calafiore. Multi-period portfolio optimization with linear control policies. *Automatica*, 44(10):2463 – 2473, 2008.

[7] William F. Shadwick Con Keating. A universal performance measure. *Journal of Performance Measurement*, 2002.

[8] George Dantzig and Gerd Infanger. Multi-stage stochastic linear programs for portfolio optimization. *Annals of Operations Research*, 45:59–76, 1993. 10.1007/BF02282041.

[9] J.J. Dongarra, Society for Industrial, and Applied Mathematics. *Linpack Users' Guide*. Miscellaneous Bks. Society for Industrial and Applied Mathematics, 1979.

[10] D. L. Donoho* and J. Tanner. Sparse nonnegative solution of underdetermined linear equations by linear programming. *Proceedings of the National Academy of Sciences*, 102:9446–9451, 2005.

[11] E. Erdogan, D. Goldfarb, and G. Iyengar. Robust portfolio management. http://citeseer.ist.psu.edu/719657.html.

[12] Peter C Fishburn. Mean-risk analysis with risk associated with below-target returns. *American Economic Review*, 67(2):116–26, 1977.

[13] Lee N. Price Frank A. Sortino. Performance measurement in a downside risk framework. *The Journal of Investing*, 3(3):59–64, 1994.

[14] M. P. Friedlander. Exact regularization of linear programs. Tech. Rep. TR-2005-31, Dept. Computer Science, University of British Columbia, Vancouver, December 2005.

[15] Michael P. Friedlander and Paul Tseng. Exact regularization of convex programs. *SIAM Journal on Optimization*, 18(4):1326–1350, 2007.

[16] Norio Hibiki. Multi-period stochastic optimization models for dynamic asset allocation. *Journal of Banking and Finance*, 30(2):365 – 390, 2006. Risk Management and Optimization in Finance.

[17] Kurt Hornik. The R FAQ, 2010. ISBN 3-900051-08-9.

[18] G. Iyengar and D. Goldfarb. Robust portfolio selection problems. http://citeseer.ist.psu.edu/goldfarb01robust.html.

[19] C. D. Mol D. Giannone J. Brodie, I. Daubechies and I. Loris. Sparse and stable markowitz portfolios, 2008.

[20] Hiroshi Konno and Hiroshi Shirakawa. Equilibrium relations in a capital asset market: A mean absolute deviation approach. *Asia-Pacific Financial Markets*, 1:21–35, 1994. 10.1007/BF02425207.

[21] Hiroshi Konno and Hiroaki Yamazaki. Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market. *Management Science*, 37(5), May 1991.

[22] John Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets: A reply. *The Review of Economics and Statistics*, 51(2):pp. 222–224, 1969.

[23] James C. T. Mao. Models of capital budgeting, e-v vs e-s. *The Journal of Financial and Quantitative Analysis*, 4(5):pp. 657–675, 1970.

[24] Harry Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, Mar 1952.

[25] Harry Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. Wiley and Sons, Inc., 1959.

[26] Rachev Martin and Siboulet. Phi-alpha optimal portfolios and extreme risk management. *Wilmott Magazine of Finance*, 2003.

[27] R.C. Merton. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, pages 1–39, December 1980.

[28] Jan Mossin. Equilibrium in a capital asset market. *Econometrica*, 34(4):pp. 768–783, 1966.

[29] JM Eber P. Artzner, F. Delbaen. Thinking coherently. *Risk*, 1997.

[30] M.J. P. Artzner, F. Delbaen and D Health. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[31] G.Ch Pflug. *Probabilistic Constrained Optimization: Methodology and Applications*, chapter Some Remarks on the Value-at-Risk and Conditional Value-at-Risk. Springer, 2000.

[32] Mustafa Pnar. Robust scenario optimization based on downside-risk measure for multi-period portfolio selection. *OR Spectrum*, 29:295–309, 2007. 10.1007/s00291-005-0023-2.

[33] James P. Quirk and Rubin Saposnik. Admissibility and measurable utility functions. *The Review of Economic Studies*, 29(2):pp. 140–146, 1962.

[34] S. Uryasev R. T. Rockafellar and M. Zabarankin. Generalized deviations in risk analysis. *Finance and Stochastics*, pages 51–74, 2006.

[35] Chambers J. M. Recker R. A. and Wilks A. R. *New S Language*. Chapman and Hall/CRC, 1988.

[36] R. Rockafellar, Stan Uryasev, and Michael Zabarankin. Generalized deviations in risk analysis. *Finance and Stochastics*, 10:51–74, 2006. 10.1007/s00780-005-0165-8.

[37] R. T. Rockafellar and S. P. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

[38] R. T. Rockafellar and S. P. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26:1443–1471, 2002.

[39] R. Terrell Rockafellar. *Convex Analysis*. Princeton University Press, 1997.

[40] A. D. Roy. Safety first and the holding of assets. *Econometrica*, 20(3):pp. 431–449, 1952.

[41] R. Schneider. *Convex Bodies: The Brunn-Minkowski Theory*. Encyclopedia of Mathematics and Its Applications. Cambridge University Press, 1993.

[42] S. I. Sergeev. Algorithms to solve some problems of concave programming with linear constraints. *Automation and Remote Control*, 68(3):399–412, 2007.

[43] William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):pp. 425–442, 1964.

[44] Van der Meer Sortino. Downside risk. *The Journal of Portfolio Management*, 17(4):27–31, 1991.

[45] Jack L. Treynor. Market value, time and risk. *Unpublished manuscript*, 1961.

[46] Jack L. Treynor. Toward a theory of market value of risky assets. *Unpublished manuscript. A final version was published in 1999, in Asset Pricing and Portfolio Performance: Models, Strategy and Performance Metrics. Robert A. Korajczyk (editor) London: Risk Books, pp. 15-22.*, 1961.

[47] F. Nogales V. DeMiguel, L. Garlappi and R. Uppal. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Forthcoming in Management Science*, 2007.

[48] L. Garlappi V. DeMiguel and R. Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy. *Review of Financial Studies*, 2007.

[49] R. Welsch. Robustrisk: Using robust methods to improve investment performance, May 2008.

VITA

Yun Zhang earned a Bachelor of Science in Mathematics from Peking University, China and a Master of Science in Applied Mathematics from the University of Washington. In 2012, she graduated with a Doctor of Philosophy at the University of Washington in Applied Mathematics.

She welcomes your comments to `zhangyun@uw.edu`.