



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

Master's Thesis

Stock Movement Prediction Leveraging News Data and Support Vector Machines

JongKyu Son

The Graduate School
Sungkyunkwan University
Department of Mathematics

Master's Thesis

Stock Movement Prediction Leveraging
News Data and Support Vector Machines

2022

JongKyu Son

Master's Thesis

Stock Movement Prediction Leveraging News Data and Support Vector Machines

JongKyu Son

The Graduate School
Sungkyunkwan University
Department of Mathematics

Stock Movement Prediction Leveraging News Data and Support Vector Machines

JongKyu Son

A Master's Thesis Submitted to the Department of Mathematics and
the Graduate School of Sungkyunkwan University in partial
fulfillment of the requirements for the degree of Master of Science in
Mathematics

april 2025

Supervised by
WooCheol Choi
Major Advisor

This certifies that the Master's Thesis
of JongKyu Son is approved

Committee Chair :

Committee Member :

Major Advisor :

The Graduate School
Sungkyunkwan University
June 2025

Table of Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 2 | Data Collection and Feature Engineering | 3 |
| 2.1 | Stock Market Data | 4 |
| 2.2 | News Data | 6 |
| 3 | Methodology | 9 |
| 3.1 | Support Vector Machine | 9 |
| 3.2 | Problem Definition | 10 |
| 3.3 | Data Construction and Preprocessing | 11 |
| 3.4 | Synthetic Minority Over-sampling Technique | 14 |
| 3.5 | Performance Metrics | 15 |
| 4 | Experimental Results and Analysis | 16 |
| 5 | Conclusion | 19 |
| | Reference | 21 |
| | Appendix | 23 |
| A | Introduction | 23 |
| B | Zeroth-Order Optimization | 24 |
| B.1 | Gaussian Smoothing | 24 |
| B.2 | Random Gradient-Free Oracles | 31 |
| B.3 | Gaussian Smoothing Gradient Method | 35 |
| C | Riemannian Optimization for Zeroth-Order Method | 38 |
| D | Numerical Examples | 44 |
| E | Reference | 46 |

Abstract

Stock Movement Prediction Leveraging News Data and Support Vector Machines

This study explores the impact of news headline sentiment and keyword frequency on short-term stock price movements by integrating them with technical indicators. Focusing on KOSPI-listed companies in the shipbuilding sector, we extracted indicators such as moving averages, RSI, MACD, and Bollinger Bands from stock data from the year 2024, and performed sentiment classification on news headlines using KoFinBERT. Keyword occurrences related to contracts, the U.S., and earnings were quantified and normalized.

Using these combined features, a Support Vector Machine (SVM) model was trained to classify next-day stock movements into two categories: significant rise (Class 1) and significant fall (Class 0). While the model showed limited performance for Class 1 due to class imbalance, it achieved relatively strong results for Class 0 (precision 0.76, recall 0.80), suggesting that sentiment and technical signals are useful in detecting stable price patterns. These findings highlight the potential of combining textual and market data for stock prediction and the need for improved methods to address data imbalance.

Keywords: News Headlines, Sentiment Analysis, Stock Price Prediction, Support Vector Machine (SVM), Technical Indicators

Chapter 1

Introduction

In the stock market, information plays a vital role in investor decision-making. In particular, unstructured text data, for instance news headlines, can reflect market sentiment and expectations, providing useful signals for stock price forecasting [1]. Recent studies have highlighted that not only the content of news but also its sentiment and frequency can influence investor behavior and market movements [2].

In Korea, market reactions to news are often asymmetric, and vary by sector, with industries like shipbuilding showing high sensitivity to external information [3]. This study focuses on the relationship between news sentiment and short-term stock trends within the Korean shipbuilding industry.

Figure 1 presents a comparison between the News Sentiment Index (NSI), published by the Bank of Korea, and the KOSPI closing price. The correlation between the two suggests that aggregate news sentiment aligns closely with market expectations.

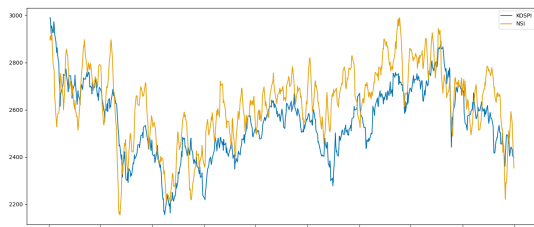


Figure 1: KOSPI and NSI

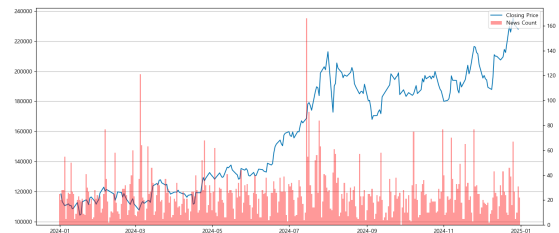


Figure 2: Closing Price and News Count

Based on the observation of the KOSPI closing price, it became evident that news may be influenced by overall market movements. To narrow the scope of analysis from the entire stock market, this study focuses on HD KSOE, a leading shipbuilding company listed on the KOSPI, as a case example. Figure 2 presents a comparison between the number of news headlines and the company’s daily closing price in 2024. While a noticeable spike in news volume occurred in mid-July the number of news articles alone was not sufficient to fully explain the price movement. This highlighted the need for a more refined and quantitative approach to analysis.

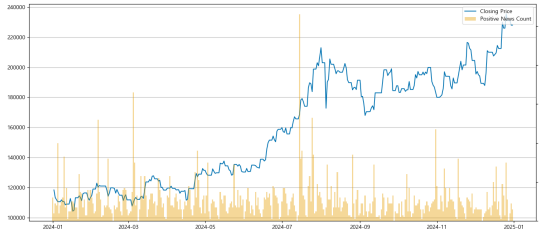


Figure 3: Closing Price and Positive News Count

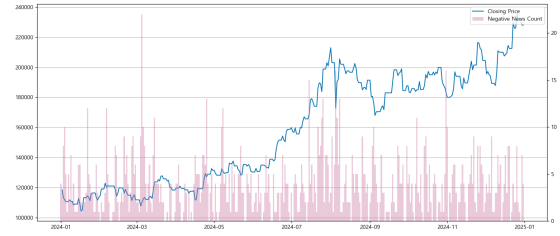


Figure 4: Closing Price and Negative News Count

To address this limitation, we apply KoFinBERT, a domain-specific sentiment analysis model trained on Korean financial news, to classify headlines as positive, neutral, or negative. As shown in Figures 3 and 4, certain time periods reveal a tendency for positive headlines to precede or accompany price increases. On the other hand, negative news appears consistently across the entire period, but its direct correlation with price decreases is weaker. These findings suggest that stock price reactions may be asymmetric with respect to news sentiment, underscoring the need for prediction models that incorporate both volume and directional sentiment. The source code for all experiments is available at <https://github.com/SonJongKyu/news-driven-stock-prediction>, allowing for full reproducibility.

Chapter 2

Data Collection and Feature Engineering

This chapter describes the types of data collected for the study and the rationale behind their selection. In 2.1 Stock Market Data, we introduce the process of crawling and preprocessing stock-related information for each company, including the closing price, price change from the previous day, opening price, high, low, and trading volume. These variables were cleaned and structured to match the requirements of the predictive modeling framework. In 2.2 News Data, we present the method for collecting news data by searching each company name and extracting news titles and URLs published in 2024. The collected data were further processed through sentiment classification and normalized using the frequency of specific keywords.

The shipbuilding industry, known for its high sensitivity to news among KOSPI-listed sectors, was selected as the target industry for this study. Given its strong dependence on overseas orders, the industry is considered particularly responsive to global economic trends and policy developments in major economies such as the United States and China [3].

The analysis period was limited to the year 2024 to ensure the inclusion of data for all 11 major shipbuilding companies and to reflect the most recent market trends and corporate responses. This temporal scope allows for a timely and consistent assessment of the relationship between news events and stock price movements.

2.1 Stock Market Data

This study collected daily stock price data for 11 major shipbuilding-related companies listed on the KOSPI. The companies were selected based on their relevance to the shipbuilding industry and their classification under related sectors. The stock price data, including daily closing price, trading volume, and rate of change, were obtained from the official source of Naver Finance (<https://finance.naver.com>).

We constructed several technical indicators using stock data crawled from the official Naver Finance website by querying each company's stock code. These indicators were selected based on their widespread use in financial analysis and their ability to capture short-term market dynamics relevant to stock price prediction [4, 5].

| Company Name | Stock Code |
|--|------------|
| HD Korea Shipbuilding & Offshore Engineering | 009540 |
| HD Hyundai Marine Solution | 443060 |
| HD Hyundai Marine Engine | 071970 |
| HD Hyundai Mipo Dockyard | 010620 |
| HD Hyundai Heavy Industries | 329180 |
| STX Engine | 077970 |
| Samsung Heavy Industries | 010140 |
| Sejin Heavy Industries | 075580 |
| Korea Carbon | 017960 |
| Hanwha Engine | 082740 |
| Hanwha Ocean | 042660 |

Table 1: Shipbuilding Companies and Stock Codes

we constructed several technical indicators using stock data crawled from the official Naver Finance website by querying each company's stock code. These indicators were selected based on their widespread use in financial analysis and their ability to capture short term market dynamics relevant to stock price prediction.

- MA5 (5-day Moving Average): Calculates the average of closing prices over the past five trading days to capture short-term trends. It is useful for identifying persistent

directional movement in stock prices and helps assess whether a news event occurs during an ongoing upward or downward trend.

- MA10 (10-day Moving Average): Calculates the average of closing prices over the past ten trading days to capture medium-term trends. It is useful for identifying persistent directional movement in stock prices and helps assess whether a news event occurs during an ongoing medium-term upward or downward trend.
- RSI (Relative Strength Index): A momentum-based indicator used to determine whether a stock is in an overbought or oversold condition. In news-based prediction, it helps detect potential reversals by signaling when the price may be due for a correction despite the sentiment of recent news.
- MACD (Moving Average Convergence Divergence): Measures the difference between short-term and long-term price trends to identify potential turning points. It is particularly useful in detecting changes that either align with or contradict the sentiment inferred from news, helping to evaluate whether a market reaction is reinforced or diverging from the psychological signal conveyed by the news.
- BB_upper (Bollinger Bands Upper Band): This indicator reflects price volatility and is used to detect whether a stock is approaching a local peak. When a touch of the upper band coincides with positive news sentiment, it may signal that the stock is already overheated, suggesting a possible price correction despite the optimistic outlook.

These indicators were incorporated into the model as features to enhance its ability to recognize short-term and mid-term technical patterns, detect trend reversals, and account for momentum or volatility conditions. Their inclusion allows the predictive framework to integrate not only external news signals but also intrinsic market behavior, thereby improving classification accuracy for short-term stock movements.

2.2 News Data

News articles were collected by querying each company on Naver, a major Korean news portal, and extracting all article titles published in the year 2024. The number of news titles collected per company is summarized as follows:

| Company Name | News Count |
|--|------------|
| HD Korea Shipbuilding & Offshore Engineering | 8,919 |
| HD Hyundai Marine Solution | 3,787 |
| HD Hyundai Marine Engine | 1,301 |
| HD Hyundai Mipo Dockyard | 3,989 |
| HD Hyundai Heavy Industries | 13,561 |
| STX Engine | 841 |
| Samsung Heavy Industries | 8,320 |
| Sejin Heavy Industries | 691 |
| Korea Carbon | 965 |
| Hanwha Engine | 4,876 |
| Hanwha Ocean | 16,126 |

Table 2: Company name and News counts

Since the collected news data consisted of unstructured text in the form of article titles, the first preprocessing step was sentiment classification. This step aimed to quantify the emotional tone of each headline using a pretrained Korean financial language model.

We used KoFinBERT, a domain-specific language model trained on financial news, to classify each headline into positive, neutral, or negative sentiment. This approach is similar to recent studies that employ contextual embeddings and sentiment classification to model the influence of news on stock behavior [6, 7].

Table 3 shows an example of the sentiment classification process using the KoFinBERT model. The input headline is first tokenized into subword units, and each token is evaluated for sentiment polarity by the model.

In this example, negative and positive cues such as “불확실성” and “미래성장” are identified, and the overall sentiment is classified as “Positive” with a confidence score of 81.00%.

| Step | Description |
|------------------|---|
| News Title | 조선·해운업계, 불확실성 줄이고 미래성장 속도낸다 |
| Tokenization | ‘조선’, ‘해운’, ‘업계’, ‘불확실성’, ‘줄이고’, ‘미래성장’, ‘속도’, ‘낸다’ |
| KoFinBERT Model | - ‘불확실성’ → -1 (negative) - ‘줄이고’ → 0 (neutral) - ‘미래성장’ → +1 (positive) - ‘속도’ → +1 (positive) |
| Sentiment | Positive |
| Confidence Score | 81.00% |

Table 3: Example of Sentiment Classification Process Using KoFinBERT

Second, we found keywords related to ‘Contract’, ‘Earnings’, and ‘USA’ in each news title. To capture topic-specific signals from news headlines, three keyword groups were manually defined based on domain knowledge of the shipbuilding industry:

- Contract-related: “order”, “contract”, “delivery”, “supply”, “export”
- Earnings-related: “earnings”, “profit”, “sales”, “surplus”, “deficit”
- U.S.-related: “United States”, “U.S.”, “Federal Reserve”, “interest rate”, “inflation”

In addition to sentiment analysis, we extracted keyword features related to contracts, earnings, and the U.S. economy. These keyword groups were selected based on their economic significance and known macroeconomic drivers in the shipbuilding industry. Specifically, contract-related terms were chosen because they directly reflect revenue-generating activities; earnings-related keywords represent key indicators of corporate performance; and U.S. all of which significantly affect the shipbuilding sector.

To numerically capture the topical relevance of these keywords within each headline, we calculated TF-IDF (Term Frequency–Inverse Document Frequency) scores. TF-IDF quantifies the relative importance of a word within a document and across the corpus, with higher scores indicating greater contribution to meaning. This approach allowed us to transform keyword occurrences into weighted feature values that reflect their contextual prominence [8].

This multi-faceted representation of news data served as a foundation for the predictive model.

Chapter 3

Methodology

3.1 Support Vector Machine

This study develops a binary classification model to predict short-term stock price movements by leveraging both news data and technical indicators from shipbuilding companies listed on the KOSPI. Accurate prediction requires algorithms capable of distinguishing between upward and downward trends, especially when combining structured indicators with unstructured news sentiment. Support Vector Machine (SVM) was selected for its strong performance in high-dimensional spaces and its ability to handle complex feature interactions.

The dataset includes 15 features, combining sentiment polarity scores, technical indicators (e.g., MA, RSI, MACD), and TF-IDF-based keyword frequencies. SVM is well-suited for mixed feature spaces and has shown robust generalization on relatively small datasets.

Prior research supports the use of SVM in financial forecasting. Zhen et al.[5] integrated investor sentiment and feature selection into an SVM model, while Fu and Zhang[8] combined sentiment analysis with technical indicators, achieving strong results in volatile markets. These studies reinforce the methodological foundation of this work.

3.2 Problem Definition

This study aims to formulate a binary classification model to predict short-term stock price movements by utilizing news data and technical indicators of major listed companies in the shipbuilding industry. The prediction target is the one-day-ahead rate of change, and to focus on economically meaningful fluctuations, we applied a threshold-based labeling scheme.

Specifically, samples showing a price increase of +4% or more on the next day were labeled as Class 1 (upward movement), while those with a decrease of -1% or more were labeled as Class 0 (downward movement). This asymmetric thresholding was chosen for two main reasons. First, in the shipbuilding sector, price surges driven by large contracts or global policy announcements tend to be abrupt and significant, whereas declines often unfold more gradually. Second, price changes within the range of -1% to +4% are more likely to reflect noise or routine market activity, rather than clear reactions to news events.

To enhance the model's ability to learn from meaningful patterns, we excluded these intermediate-range samples from the training set. By doing so, the classifier focuses only on high-confidence signals, reducing the risk of confusion from ambiguous cases.

This thresholding strategy is consistent with the approach of Jin et al.[2], who framed stock prediction as linking sentiment-labeled news to directional price changes. It also aligns with Cho and Cho[3], who emphasized the importance of isolating event-driven effects in financial modeling. Together, these considerations support a classification framework centered on capturing clear and economically relevant market responses.

3.3 Data Construction and Preprocessing

Two types of data were utilized in this analysis:

- Technical indicator data: Includes features such as closing price, Trading volume, moving averages (MA), MACD, RSI, and Bollinger Bands.

| Company | Date | price | volume | MA5 | MA10 | RSI | MACD | BB_upper |
|--------------|------------|--------|---------|--------|--------|-------|---------|-----------|
| HD KSOE | 2024-01-29 | 113500 | 112906 | 114160 | 111750 | 53.83 | -541.69 | 118774.84 |
| HD KSOE | 2024-01-30 | 111700 | 139768 | 113800 | 111990 | 51.05 | -600.28 | 117683.09 |
| HD KSOE | 2024-01-31 | 113500 | 133430 | 114220 | 112900 | 56.50 | -495.76 | 117683.09 |
| HD KSOE | 2024-02-01 | 114700 | 160045 | 113980 | 113860 | 57.29 | -312.49 | 117981.76 |
| HD KSOE | 2024-02-02 | 119100 | 315586 | 114500 | 114440 | 58.06 | 185.66 | 119206.52 |
| HD KSOE | 2024-02-05 | 123000 | 302387 | 116400 | 115280 | 67.34 | 884.93 | 121276.67 |
| HD KSOE | 2024-02-06 | 120400 | 188255 | 118140 | 115970 | 71.51 | 1215.31 | 122380.13 |
| HD KSOE | 2024-02-07 | 121400 | 152311 | 119720 | 116970 | 71.73 | 1540.08 | 123552.43 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| STX Engine | 2024-06-05 | 14190 | 127141 | 13756 | 13573 | 43.62 | 53.94 | 14715.09 |
| STX Engine | 2024-06-07 | 14350 | 126858 | 13996 | 13680 | 58.25 | 100.15 | 14784.04 |
| STX Engine | 2024-06-10 | 15380 | 333380 | 14414 | 13853 | 70.22 | 217.37 | 15080.16 |
| STX Engine | 2024-06-11 | 14980 | 138864 | 14584 | 13998 | 67.22 | 274.83 | 15224.66 |
| STX Engine | 2024-06-12 | 15700 | 289616 | 14920 | 14238 | 77.85 | 374.15 | 15492.27 |
| STX Engine | 2024-06-13 | 15580 | 150963 | 15198 | 14477 | 75.22 | 438.12 | 15694.36 |
| STX Engine | 2024-06-14 | 15780 | 406239 | 15484 | 14740 | 74.26 | 499.21 | 15906.24 |
| STX Engine | 2024-06-17 | 15980 | 132619 | 15604 | 15009 | 77.40 | 557.33 | 16194.48 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Hanwha Ocean | 2024-12-11 | 33100 | 2889057 | 31840 | 33165 | 42.03 | -359.31 | 40684.12 |
| Hanwha Ocean | 2024-12-12 | 33150 | 3246708 | 31960 | 32880 | 33.33 | -346.05 | 40141.15 |
| Hanwha Ocean | 2024-12-13 | 34050 | 2956401 | 32390 | 32785 | 38.06 | -259.92 | 39529.43 |
| Hanwha Ocean | 2024-12-16 | 33700 | 1963347 | 33160 | 32835 | 41.84 | -217.40 | 39063.90 |
| Hanwha Ocean | 2024-12-17 | 32300 | 2384863 | 33260 | 32615 | 36.42 | -293.29 | 38775.26 |
| Hanwha Ocean | 2024-12-18 | 32350 | 4303943 | 33110 | 32475 | 37.63 | -345.41 | 38326.02 |
| Hanwha Ocean | 2024-12-19 | 32250 | 3341661 | 32930 | 32445 | 40.07 | -390.29 | 38133.10 |
| Hanwha Ocean | 2024-12-20 | 33400 | 6391430 | 32800 | 32595 | 50.76 | -329.27 | 37393.34 |

Table 4: Technical indicator data

- News-based feature data: including the number of news articles, sentiment classification results (positive, negative), and TF-IDF-based keyword frequencies related to terms such as "contract," "earnings," and "United States."

| Date | Company | Positive Ratio | Negative Ratio | Contract | Earnings | US |
|------------|-----------------|----------------|----------------|----------|----------|-----|
| 2024-01-01 | HD KSOE | 0.6190 | 0.3810 | 0.0214 | 0 | 0 |
| 2024-01-01 | HD MS | 0.7500 | 0.2500 | 0 | 0 | 0 |
| 2024-01-01 | HD ME | 1.0000 | 0 | 0 | 0 | 0 |
| 2024-01-01 | HD Hyundai Mipo | 0.8333 | 0.1667 | 0 | 0 | 0 |
| 2024-01-01 | HD HHI | 0.7826 | 0.2174 | 0.0088 | 0 | 0 |
| 2024-01-01 | STX Engine | 0 | 1.0000 | 0 | 0 | 0 |
| 2024-01-01 | SHI | 0.7143 | 0.2857 | 0.0103 | 0 | 0 |
| 2024-01-01 | Sejin Heavy | 0.6667 | 0.3333 | 0 | 0 | 0 |
| 2024-01-01 | Korea Carbon | 0.5000 | 0.5000 | 0 | 0 | 0 |
| 2024-01-01 | Hanwha Engine | 0.4000 | 0.6000 | 0.0827 | 0 | 0 |
| 2024-01-01 | Hanwha Ocean | 0.6400 | 0.3600 | 0.0076 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... |
| 2024-06-17 | HD KSOE | 0.7600 | 0.2400 | 0.0405 | 0 | 0 |
| 2024-06-17 | HD MS | 0.9000 | 0.1000 | 0 | 0.0186 | 0 |
| 2024-06-17 | HD ME | 1.0000 | 0 | 0 | 0 | 0 |
| 2024-06-17 | HD Hyundai Mipo | 0.7500 | 0.2500 | 0.2343 | 0 | 0 |
| 2024-06-17 | HD HHI | 0.8293 | 0.1707 | 0.1531 | 0 | 0 |
| 2024-06-17 | SHI | 0.7895 | 0.2105 | 0.1346 | 0.0095 | 0 |
| 2024-06-17 | Sejin Heavy | 1.0000 | 0 | 0.3239 | 0 | 0 |
| 2024-06-17 | Korea Carbon | 0.6667 | 0.3333 | 0 | 0 | 0 |
| 2024-06-17 | Hanwha Engine | 0.7143 | 0.2857 | 0 | 0 | 0 |
| 2024-06-17 | Hanwha Ocean | 0.7188 | 0.2813 | 0.0711 | 0.0212 | 0 |
| ... | ... | ... | ... | ... | ... | ... |
| 2024-12-31 | HD KSOE | 0.5882 | 0.4118 | 0 | 0.0332 | 0 |
| 2024-12-31 | HD MS | 0.6667 | 0.3333 | 0 | 0 | 0 |
| 2024-12-31 | HD Hyundai Mipo | 0.8182 | 0.1818 | 0 | 0 | 0 |
| 2024-12-31 | HD HHI | 0.7368 | 0.2632 | 0.0305 | 0.0132 | 0 |
| 2024-12-31 | SHI | 0.6471 | 0.3529 | 0.0229 | 0 | 0 |
| 2024-12-31 | Hanwha Engine | 0.8000 | 0.2000 | 0 | 0 | 0 |
| 2024-12-31 | Hanwha Ocean | 0.8148 | 0.1852 | 0.0236 | 0 | 0 |

Table 5: News-based feature data

The two datasets were merged based on date and company name. Missing values generated during the merging process were imputed with zeros. From the combined dataset, a total of 15 key features encompassing both technical and textual information were selected for model training.

To eliminate differences in feature scale, Z-score normalization was performed using `StandardScaler`. Given the temporal nature of the dataset, we employed a time-based partitioning strategy: the data was sorted chronologically and split into 70% for training and 30% for testing. This method helps prevent data leakage from future information and allows the model to be evaluated on truly unseen, temporally subsequent data—closely simulating real-world forecasting scenarios.

The choice of a 30% testing split balances two key considerations. First, it secures a sufficiently large training set to enable the model to learn meaningful patterns from historical data. Second, it retains a significant testing period that captures diverse market conditions—including potential regime shifts, volatility clusters, and sentiment changes—which are crucial for robust evaluation in financial contexts. Using a non-trivial portion of temporally future data simulates realistic forecasting scenarios and enhances the reliability of performance metrics such as accuracy, precision, and recall.

This design follows common best practices in financial machine learning, where time-aware validation is critical for developing deployable models [?, ?]. By testing on a future window that is representative of actual use cases, the model’s predictive power can be more accurately assessed.

The process of preprocessing news data—through sentiment classification, TF-IDF transformation, and normalization—aligns with recent best practices in explainable financial AI. Sun and Li [9] proposed a similar preprocessing pipeline that captures intraday and post-market sentiment to enhance the transparency of prediction models. Furthermore, Leone et al. [10] illustrated the importance of sentiment refinement using ESG news, underscoring the value of careful feature engineering when integrating unstructured textual inputs.

3.4 Synthetic Minority Over-sampling Technique

In the original dataset used to train the model, a significant class imbalance was observed between the two target classes. As can be seen in the table 6, the number of samples classified as class 1 (indicating an upward movement in stock price) (183) was significantly lower than class 0 (536), indicating a downward or insignificant movement. This imbalance hinders the ability of the classifier to correctly identify the minority class, resulting in high precision but low recall for class 1.

| Class | Before SMOTE | After SMOTE |
|---------|--------------|-------------|
| Class 0 | 536 | 536 |
| Class 1 | 183 | 536 |

Table 6: Class distribution before and after SMOTE

To address this issue, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to the training dataset. SMOTE is a data augmentation technique that generates synthetic samples of the minority class by interpolating between existing examples and their k -nearest neighbors in the feature space. In this study, the number of neighbors was set to $k = 3$ to generate diverse and reliable synthetic data points.

As a result of applying SMOTE, the class distribution was balanced, with both Class 0 and Class 1 consisting of 536 instances, as reflected in Table 6. This balanced training set allowed the Support Vector Machine (SVM) model—particularly sensitive to data distribution and margin optimization—to better learn discriminative patterns from both classes. After resampling, the model was trained using a linear kernel SVM and subsequently evaluated on the original (imbalanced) test set to ensure realistic generalization performance.

Although the primary focus of Zhen et al. [5] was on sentiment and multi-feature fusion, their work also highlights the challenges posed by class imbalance in stock prediction. Their study supports the need for data balancing techniques such as SMOTE when building robust classifiers in financial applications.

3.5 Performance Metrics

In addition to accuracy, the performance of the classification model was evaluated using the `classification_report()` function provided by the Scikit-learn library. This function generates a detailed summary of the model's performance across each class, including the metrics of precision, recall, F1-score, and support. These metrics are especially useful for evaluating models trained on imbalanced datasets, such as those commonly found in financial forecasting.

- Precision: The proportion of predicted positive cases that were actually positive. It is defined as $\text{Precision} = \frac{TP}{TP+FP}$.
- Recall: The proportion of actual positive cases that were correctly identified by the model. It is given by $\text{Recall} = \frac{TP}{TP+FN}$.
- F1-score: The harmonic mean of precision and recall, calculated as $F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$. This metric provides a balanced measure of a model's performance, particularly under class imbalance.

| Actual \ Predicted | Positive (1) | Negative (0) |
|--------------------|---------------------|---------------------|
| Positive (1) | TP (True Positive) | FN (False Negative) |
| Negative (0) | FP (False Positive) | TN (True Negative) |

Table 7: Confusion Matrix Terminology

These metrics were selected to offer a more nuanced evaluation of the classification model, particularly in the context of financial data where minority class prediction is critical.

The choice of evaluation metrics is consistent with prior studies such as Srinivas et al. [7] and Sun and Li [9], which emphasize precision, recall, and F1-score when dealing with imbalanced financial datasets. These metrics provide a more informative assessment than accuracy alone, especially when minority classes carry greater financial relevance.

Chapter 4

Experimental Results and Analysis

The performance of the model was evaluated based on the test dataset using classification metrics. The results are as follows:

| | Precision | Recall | F1-score | Support |
|---------------------|-----------|--------|----------|---------|
| 0 | 0.76 | 0.80 | 0.78 | 234 |
| 1 | 0.26 | 0.21 | 0.23 | 75 |
| Accuracy | | | 0.66 | 309 |
| Macro avg | 0.51 | 0.51 | 0.51 | 309 |
| Weighted avg | 0.64 | 0.66 | 0.65 | 309 |

Table 8: Classification Report for SVM Model

The classification report reveals a notable disparity in performance between the two classes. For Class 0, which represents downward stock movements, the model achieved a precision of 0.76, recall of 0.80, and an F1-score of 0.78. These results indicate that the model is capable of identifying downward movements with a high degree of accuracy and consistency. In contrast, Class 1, corresponding to upward movements, yielded significantly lower performance, with a precision of 0.26, recall of 0.21, and F1-score of 0.23.

These results indicate that the model performed better in predicting the downward movement class (Class 0), while it struggled with the upward movement class

(Class 1).

There are two main reasons that may explain the relatively higher performance for Class 0. First, negative news tends to contain clearer sentiment expressions. News related to stock declines often includes explicitly negative language (e.g., "loss," "delay," "risk"), which the model may find easier to learn and classify. Second, the model's performance appears to have been affected by class imbalance. Although SMOTE was applied to balance the training dataset, Class 1 remained underrepresented in the test set, which may have limited the model's ability to generalize well for that class.

To better understand the contribution of each input feature to the classification outcome, we extracted the feature importance values from the trained linear SVM model. Since the linear SVM defines a decision boundary using a linear combination of feature weights, the absolute value of each coefficient can be interpreted as the relative importance of the corresponding feature. Figure 5 illustrates the feature importance based on the absolute values of the learned coefficients.

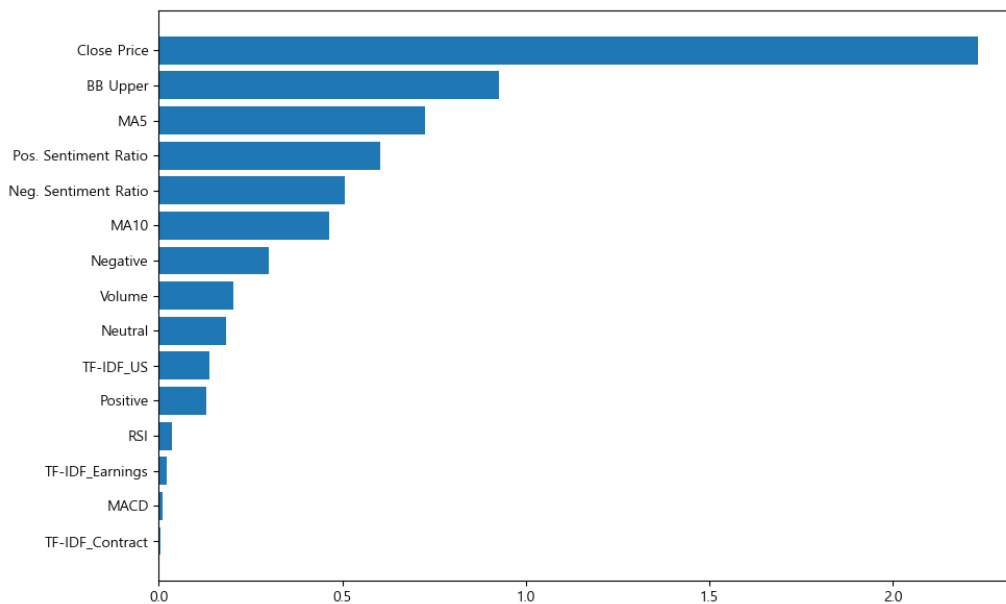


Figure 5: Feature Importance from Linear SVM Coefficients

As shown in the figure titled "Feature Importance from Linear SVM Coefficients", the most influential feature was Close price, followed by Bollinger Band upper bound (BB Upper) and moving averages (MA5 and MA10). Among the news-based variables, positive and negative sentiment ratios and TF-IDF scores related to the U.S. and contract keywords demonstrated moderate impact, although they were generally less influential than technical indicators. This result suggests that historical price patterns had a more dominant role in shaping the model's predictions compared to text-derived features in this setting.

The interpretability of sentiment features remains an important consideration. Leone et al. [10] examined how sentiment scores—particularly those related to ESG factors—can influence model decision-making, even if their relative weight is lower than that of numerical indicators. This points to the potential of improving sentiment signal quality or interpretability through domain adaptation or feature refinement in future studies. Additionally, Lee et al. [11] demonstrated that integrating features derived from large language models (LLMs) improved classification performance, suggesting a potential direction for expanding this framework in future work.

Chapter 5

Conclusion

This study aimed to explore the predictive power of financial news and technical indicators on short-term stock price movements, using companies in the shipbuilding sector listed on the KOSPI. By combining structured market data with sentiment-annotated news headlines, we constructed a binary classification model that predicts significant price surges or drops on the following trading day. A linear Support Vector Machine (SVM) was employed to strike a balance between interpretability and performance.

The results show that the model performed better in identifying downward price movements (Class 0), with relatively lower predictive power for upward movements (Class 1). Feature importance analysis revealed that traditional technical indicators had a more pronounced impact on the model's decision boundary than news-derived features such as sentiment polarity or TF-IDF scores. Nonetheless, sentiment-based features provided complementary value, especially in highlighting market optimism or pessimism around specific time periods.

These findings suggest that investor reactions to news may be asymmetrical, and that negative sentiment tends to manifest more clearly in stock behavior. Moreover, the persistence of class imbalance in the test set, despite SMOTE adjustment, highlights the challenges of modeling minority-class events such as sharp price increases.

Several limitations should be noted. First, the reliance on news headlines alone

excludes the full textual richness of the articles, potentially omitting important contextual clues. Second, the use of static TF-IDF scores may limit the model’s sensitivity to nuanced language changes. Lastly, the analysis focuses on a single year and a specific industry, which may constrain the generalizability of the findings.

Future research could extend this work by incorporating deep learning models such as LSTMs or Transformers to better capture sequential dependencies in news and market data [9, 11]. The integration of large language models (LLMs), as explored by Lee et al. [11], may also improve feature richness and interpretability in sentiment-based forecasting. Additionally, combining sentiment signals from multiple sources—such as news, ESG disclosures, and macroeconomic reports—has shown promise in increasing model robustness [8]. Leone et al. [10] also emphasize the value of including ESG and global policy signals as part of the feature space, especially for industries highly sensitive to geopolitical and environmental developments.

Ultimately, this study demonstrates both the potential and current limitations of integrating textual sentiment analysis with quantitative market data. While news-based sentiment adds contextual depth, its predictive influence remains secondary to historical technical patterns. Nevertheless, a multi-source, multi-modal feature design holds promise for enhancing predictive accuracy and generalization in future financial forecasting models.

References

- [1] Boudoukh, J., Feldman, R., Kogan, S., and Richardson, M. P.: Which News Moves Stock Prices? A Textual Analysis. *NBER Working Paper No. w18725*, National Bureau of Economic Research, (2013).
- [2] Jin, Z., Ma, M., Zhou, Z., Gan, S., and others: Correlation Between News and Stock Price Based on Stock Market Indices: Can News Classification Be Used as a Tool to Make Better Decisions? *Advances in Economics Management and Political Sciences*, 82(1), 131–141 (2024).
- [3] Cho, H., and Cho, H.: *Study on Stock Price Reaction to News*. Journal of Accounting and Information, vol.24, no.1, pp.75–98 (2024).
- [4] Gao, H., and Georgopoulos, G.: Stock Price Prediction Using Sentiment and Technical Analysis. In *Proceedings of the 2024 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pp. 904–911, Toronto, Canada, (2024).
- [5] Zhen, K., Xie, D., and Hu, X.: A multi-feature selection fused with investor sentiment for stock price prediction. *Expert Systems with Applications*, 278, 127381 (2025).
- [6] Ren, Y., Liao, F., and Gong, Y.: Impact of News on the Trend of Stock Price Change: an Analysis based on the Deep Bidirectional LSTM Model. *Procedia Computer Science*, 174, 128–140 (2020).
- [7] Srinivas, S., Gadela, R., Sabu, R., Das, A., Nath, G., and Datla, V.: Effects of Daily News Sentiment on Stock Price Forecasting. *arXiv preprint*, arXiv:2308.08549 [q-fin.ST], (2023).

- [8] Fu, K., and Zhang, Y.: Incorporating Multi-Source Market Sentiment and Price Data for Stock Price Prediction. *Mathematics*, 12(10), 1572 (2024).
- [9] Sun, G., and Li, Y.: Intraday and Post-Market Investor Sentiment for Stock Price Prediction: A Deep Learning Framework with Explainability and Quantitative Trading Strategy. *Systems*, 13(5), 390 (2025).
- [10] Leone, F., Marazzina, D., and Rosamilia, N.: What's news with you: Price forecasting with global and ESG sentiment scores. *Finance Research Open*, 1(2), 100013 (2025).
- [11] Lee, Y., Kim, Y., Kim, S., and Lee, Y.: Integrating LLM-Generated Views into Mean-Variance Optimization Using the Black-Litterman Model. arXiv preprint arXiv:2504.14345 (2025).

Appendix

Zeroth-Order Optimization on Riemannian Manifolds via Gaussian Smoothing

This appendix addresses a topic in Riemannian zeroth-order optimization, which is unrelated to the main subject of this thesis but included here as a supplementary mathematical study by the author.

A Introduction

In this paper, we investigate the problem of zeroth-order optimization for both smooth and nonsmooth objective functions. In particular, we consider the following unconstrained optimization task:

$$\min_{x \in \mathbb{R}^n} f(x) \tag{1}$$

where the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ may be nonconvex but is assumed to exhibit a certain level of smoothness.

Our focus lies on the function class $C^{(1,1)}$, which comprises continuously differentiable functions with Lipschitz continuous gradients. A function $f \in C^{(1,1)}$ satisfies the condition:

$$\|\nabla f(x) - \nabla f(y)\| \leq L_1(f)\|x - y\|, \quad \forall x, y \in \mathbb{R}^n \tag{2}$$

where $L_1(f)$ denotes the Lipschitz constant of the gradient. This assumption implies the second-order upper bound:

$$|f(y) - f(x) - \langle \nabla f(x), y - x \rangle| \leq \frac{L_1(f)}{2} \|y - x\|^2, \quad \forall x, y \in \mathbb{R}^n \quad (3)$$

Furthermore, if the function f satisfies the inequality

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L_1(f)}{2} \|y - x\|^2 \quad (4)$$

for all $x, y \in \mathbb{R}^n$, then f is said to be strongly convex with parameter $L_1(f)$.

Such functions are commonly encountered in machine learning and optimization settings, particularly in scenarios where gradient information is inaccessible or expensive to compute. To address these challenges, we propose zeroth-order optimization algorithms based on Gaussian smoothing and provide theoretical analysis of their convergence guarantees.

B Zeroth-Order Optimization

B.1 Gaussian Smoothing

We study zeroth-order optimization in the context of smooth and convex analysis by applying Gaussian smoothing to the objective function $f : \mathbb{R}^n \rightarrow \mathbb{R}$. The smoothed version f_μ is defined via convolution with a Gaussian kernel:

$$f_\mu(x) \stackrel{\text{def}}{=} \frac{1}{\kappa} \int_{\mathbb{R}^n} f(x + \mu u) e^{-\frac{1}{2} \|u\|^2} du \quad (5)$$

where the normalization constant is given by

$$\kappa \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} e^{-\frac{1}{2} \|u\|^2} du = \frac{(2\pi)^{\frac{n}{2}}}{[\det B]^{\frac{1}{2}}} \quad (6)$$

For any $\mu > 0$, the smoothed function f_μ is differentiable, with $\mu \geq 0$ serving

as a smoothing parameter. Notably, since $\frac{1}{\kappa} \int_{\mathbb{R}^n} u e^{-\frac{1}{2}\|u\|^2} du = 0$ it follows that for a convex function f with subgradient $g \in \partial f(x)$,

$$f_\mu(x) \geq \frac{1}{\kappa} \int_{\mathbb{R}^n} [f(x) + \mu \langle g, u \rangle] e^{-\frac{1}{2}\|u\|^2} du = f(x) \quad (7)$$

In general, the smoothed function f_μ inherits and often improves upon the analytical properties of f , particularly when $\mu > 0$. For example:

- If f is convex, then so is f_μ
- If $f \in C^{(1,1)}$, then $f_\mu \in C^{(1,1)}$ and $L_1(f_\mu) \leq L_1(f)$, since

$$\begin{aligned} \|\nabla f_\mu(x) - \nabla f_\mu(y)\| &\leq \frac{1}{\kappa} \int_{\mathbb{R}^n} \|\nabla f(x + \mu u) - \nabla f(y + \mu u)\| e^{-\frac{1}{2}\|u\|^2} du \\ &\leq L_1(f) \|x - y\|, \quad \forall x, y \in \mathbb{R}^n. \end{aligned} \quad (8)$$

From the definition in (6), we also obtain the identity

$$\ln \int_{\mathbb{R}^n} e^{-\frac{1}{2}\langle Bu, u \rangle} du = \frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln \det B.$$

Differentiating this identity with respect to B yields

$$\frac{1}{\kappa} \int_{\mathbb{R}^n} uu^T e^{-\frac{1}{2}\|u\|^2} du = B^{-1}. \quad (9)$$

Taking the trace of both sides after multiplying by B , we obtain

$$\frac{1}{\kappa} \int_{\mathbb{R}^n} \|u\|^2 e^{-\frac{1}{2}\|u\|^2} du = n. \quad (10)$$

We frequently require bounds for the moments defined by $M_p \stackrel{\text{def}}{=} \frac{1}{\kappa} \int_{\mathbb{R}^n} \|u\|^p e^{-\frac{1}{2}\|u\|^2} du$. For the most common cases, we have exact evaluations:

$$M_0 \stackrel{(6)}{=} 1, \quad M_2 \stackrel{(10)}{=} n \quad (11)$$

For other cases, we will apply the following straightforward bounds.

Lemma B.1. *Let $p \geq 0$. Then the following hold: For $p \in [0, 2]$,*

$$M_p \leq n^{\frac{p}{2}} \quad (12)$$

For $p \geq 2$, we have the two-sided bound

$$n^{\frac{p}{2}} \leq M_p \leq (p+n)^{\frac{p}{2}} \quad (13)$$

Proof. Define $\psi(p) := \ln M_p$. Let $p = (1 - \alpha) \cdot 0 + \alpha \cdot 2$, so that $\alpha = \frac{p}{2} \in [0, 1]$ for $p \in [0, 2]$. By convexity of ψ , we obtain:

$$\psi(p) \leq (1 - \alpha)\psi(0) + \alpha\psi(2) \stackrel{(10)}{=} \frac{p}{2} \ln n$$

which implies $M_p \leq n^{\frac{p}{2}}$, establishing the bound in (12). For $p \geq 2$, note that $\alpha \geq 1$, and thus $\alpha\psi(2) \leq \psi(p)$, which gives the lower bound $M_p \geq n^{\frac{p}{2}}$. To derive the upper bound for $p \geq 2$, fix any $\tau \in (0, 1)$. For all $t \geq 0$, the following inequality holds:

$$t^p e^{-\frac{\tau}{2}t^2} \leq \left(\frac{p}{\tau e}\right)^{\frac{p}{2}} \quad (14)$$

Applying this to the moment integral, we write:

$$\begin{aligned} M_p &= \frac{1}{\kappa} \int_{\mathbb{R}^n} \|u\|^p e^{-\frac{1}{2}\|u\|^2} du = \frac{1}{\kappa} \int_{\mathbb{R}^n} \|u\|^p e^{-\frac{1-\tau}{2}\|u\|^2} du \\ &\stackrel{(14)}{\leq} \frac{1}{\kappa} \left(\frac{p}{\tau e}\right)^{\frac{p}{2}} \int_{\mathbb{R}^n} e^{-\frac{1-\tau}{2}\|u\|^2} du = \left(\frac{p}{\tau e}\right)^{\frac{p}{2}} \frac{1}{(1-\tau)^{\frac{n}{2}}} \end{aligned}$$

Minimizing the right-hand side with respect to $\tau \in (0, 1)$, the minimum is achieved at $\tau = \frac{p}{p+n}$. Substituting this value yields:

$$M_p \leq \left(\frac{p}{e}\right)^{\frac{p}{2}} \left(1 + \frac{n}{p}\right)^{\frac{p}{2}} \left(1 + \frac{p}{n}\right)^{\frac{n}{2}} \leq (p+n)^{\frac{p}{2}} \quad \blacksquare$$

We now establish the following approximation bound for the smoothed function.

Theorem B.1. *Let $f \in C^{(1,1)}$. Then, for all $x \in \mathbb{R}^n$,*

$$|f_\mu(x) - f(x)| \leq \frac{\mu^2}{2} L_1(f) n. \quad (15)$$

Proof. Recall that the smoothed function is defined as

$$f_\mu(x) \stackrel{\text{def}}{=} \frac{1}{\kappa} \int_{\mathbb{R}^n} f(x + \mu u) e^{-\frac{1}{2}\|u\|^2} \mathrm{d}u,$$

while

$$f(x) = \frac{1}{\kappa} \int_{\mathbb{R}^n} f(x) e^{-\frac{1}{2}\|u\|^2} \mathrm{d}u.$$

If f is differentiable at x , we may write

$$f_\mu(x) - f(x) = \frac{1}{\kappa} \int_{\mathbb{R}^n} [f(x + \mu u) - f(x) - \mu \langle \nabla f(x), u \rangle] e^{-\frac{1}{2}\|u\|^2} \mathrm{d}u.$$

By the $C^{(1,1)}$ assumption (Lipschitz continuity of the gradient), we have

$$|f(x + \mu u) - f(x) - \mu \langle \nabla f(x), u \rangle| \leq \frac{1}{2} L_1(f) \|\mu u\|^2 = \frac{1}{2} \mu^2 L_1(f) \|u\|^2.$$

Therefore,

$$\begin{aligned} |f_\mu(x) - f(x)| &\leq \frac{1}{\kappa} \int_{\mathbb{R}^n} \frac{1}{2} \mu^2 L_1(f) \|u\|^2 e^{-\frac{1}{2}\|u\|^2} \mathrm{d}u \\ &\leq \frac{\mu^2 L_1(f)}{2} \cdot \frac{1}{\kappa} \int_{\mathbb{R}^n} \|u\|^2 e^{-\frac{1}{2}\|u\|^2} \mathrm{d}u \\ &= \frac{\mu^2 L_1(f)}{2} \cdot M_2 \stackrel{(10)}{=} \frac{\mu^2 L_1(f)}{2} \cdot n. \end{aligned}$$

■

For any positive μ , the smoothed function f_μ is differentiable. We now derive a convenient expression for its gradient. To this end, we rewrite the definition (5) by introducing a change of variables $y = x + \mu u$, which yields:

$$f_\mu(x) = \frac{1}{\mu^n \kappa} \int_{\mathbb{R}^n} f(y) e^{-\frac{1}{2\mu^2} \|y-x\|^2} dy.$$

Since both the integrand and its partial derivatives with respect to x are continuous in (x, y) , we may apply differentiation under the integral sign to obtain the gradient:

$$\nabla f_\mu(x) = \frac{1}{\mu^{n+2} \kappa} \int_{\mathbb{R}^n} f(y) e^{-\frac{1}{2\mu^2} \|y-x\|^2} B(y-x) dy.$$

Returning to the original variable $u = \frac{y-x}{\mu}$, we obtain equivalent expressions:

$$\begin{aligned} \nabla f_\mu(x) &= \frac{1}{\mu^{n+2} \kappa} \int_{\mathbb{R}^n} f(y) e^{-\frac{1}{2\mu^2} \|y-x\|^2} B(y-x) dy \\ &= \frac{1}{\mu \kappa} \int_{\mathbb{R}^n} f(x + \mu u) e^{-\frac{1}{2} \|u\|^2} B u du \\ &= \frac{1}{\kappa} \int_{\mathbb{R}^n} \frac{f(x + \mu u) - f(x)}{\mu} e^{-\frac{1}{2} \|u\|^2} B u du \\ &= \frac{1}{\kappa} \int_{\mathbb{R}^n} \frac{f(x) - f(x - \mu u)}{\mu} e^{-\frac{1}{2} \|u\|^2} B u du \\ &= \frac{1}{\kappa} \int_{\mathbb{R}^n} \frac{f(x + \mu u) - f(x - \mu u)}{2\mu} e^{-\frac{1}{2} \|u\|^2} B u du \end{aligned} \tag{16}$$

Among these, the final symmetric form (16) is particularly useful, especially in practical algorithms such as zeroth-order gradient descent. Interestingly, the gradient ∇f_μ is Lipschitz continuous even when f itself is not differentiable. This enhanced regularity is a direct consequence of Gaussian smoothing.

Lemma B.2. *Let $f \in C^{(1,1)}$. Then, the gradient of the smoothed function f_μ satisfies the following bound:*

$$\|\nabla f_\mu(x) - \nabla f(x)\| \leq \frac{\mu}{2} L_1(f)(n+3)^{3/2}, \quad \forall x \in \mathbb{R}^n. \quad (17)$$

Proof. Since $f \in C^{(1,1)}$, we can bound the deviation of gradients using the integral form of $\nabla f_\mu(x)$:

$$\begin{aligned} \|\nabla f_\mu(x) - \nabla f(x)\| &= \left\| \frac{1}{\kappa} \int_{\mathbb{R}^n} \left(\frac{f(x + \mu u) - f(x)}{\mu} - \langle \nabla f(x), u \rangle \right) B u e^{-\frac{1}{2}\|u\|^2} du \right\| \\ &\leq \frac{1}{\kappa \mu} \int_{\mathbb{R}^n} |f(x + \mu u) - f(x) - \mu \langle \nabla f(x), u \rangle| \|u\| e^{-\frac{1}{2}\|u\|^2} du. \end{aligned}$$

By the second-order upper bound property of $C^{(1,1)}$ functions (Equation (3)), we have:

$$|f(x + \mu u) - f(x) - \mu \langle \nabla f(x), u \rangle| \leq \frac{1}{2} L_1(f) \mu^2 \|u\|^2.$$

Substituting this into the integral:

$$\|\nabla f_\mu(x) - \nabla f(x)\| \leq \frac{\mu L_1(f)}{2\kappa} \int_{\mathbb{R}^n} \|u\|^3 e^{-\frac{1}{2}\|u\|^2} du = \frac{\mu}{2} L_1(f) \cdot M_3.$$

From the moment bound (Equation (13)) for $p = 3$, we know that:

$$M_3 \leq (n+3)^{3/2}.$$

Thus, the final bound becomes:

$$\|\nabla f_\mu(x) - \nabla f(x)\| \leq \frac{\mu}{2} L_1(f)(n+3)^{3/2}.$$

■

Finally, we establish one more relationship between the gradients of f and f_μ .

Lemma B.3. Let $f \in C^{(1,1)}$. Then, for any $x \in \mathbb{R}^n$, the gradient of f is bounded by the smoothed gradient as follows:

$$\|\nabla f(x)\|^2 \leq 2\|\nabla f_\mu(x)\|^2 + \frac{\mu^2}{2} L_1^2(f)(n+6)^3. \quad (18)$$

Proof. We begin by decomposing $\|\nabla f(x)\|^2$ in terms of the smoothed gradient:

$$\begin{aligned} \|\nabla f(x)\|^2 &= \left\| \frac{1}{\kappa} \int_{\mathbb{R}^n} \langle \nabla f(x), u \rangle B u e^{-\frac{1}{2}\|u\|^2} du \right\|^2 \\ &= \left\| \frac{1}{\kappa\mu} \int_{\mathbb{R}^n} [f(x + \mu u) - f(x) - (f(x + \mu u) - f(x) - \mu \langle \nabla f(x), u \rangle)] B u e^{-\frac{1}{2}\|u\|^2} du \right\|^2 \\ &= \left\| \frac{1}{\kappa} \int_{\mathbb{R}^n} \frac{f(x + \mu u) - f(x)}{\mu} B u e^{-\frac{1}{2}\|u\|^2} du \right. \\ &\quad \left. - \frac{1}{\kappa\mu} \int_{\mathbb{R}^n} [f(x + \mu u) - f(x) - \mu \langle \nabla f(x), u \rangle] B u e^{-\frac{1}{2}\|u\|^2} du \right\|^2 \end{aligned}$$

Using the inequality $\|a - b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$, we obtain:

$$\begin{aligned} \|\nabla f(x)\|^2 &\leq 2 \left\| \frac{1}{\kappa} \int_{\mathbb{R}^n} \frac{f(x + \mu u) - f(x)}{\mu} B u e^{-\frac{1}{2}\|u\|^2} du \right\|^2 \\ &\quad + \frac{2}{\mu^2} \left\| \frac{1}{\kappa} \int_{\mathbb{R}^n} [f(x + \mu u) - f(x) - \mu \langle \nabla f(x), u \rangle] B u e^{-\frac{1}{2}\|u\|^2} du \right\|^2 \end{aligned}$$

The first term on the right is simply $2\|\nabla f_\mu(x)\|^2$ by Equation (16). For the second term, we apply inequality:

$$\leq 2\|\nabla f_\mu(x)\|^2 + \frac{2}{\mu^2\kappa} \int_{\mathbb{R}^n} [f(x + \mu u) - f(x) - \mu \langle \nabla f(x), u \rangle]^2 \|u\|^2 e^{-\frac{1}{2}\|u\|^2} du.$$

Using the second-order smoothness bound (Equation (3)),

$$|f(x + \mu u) - f(x) - \mu \langle \nabla f(x), u \rangle| \leq \frac{1}{2} L_1(f) \mu^2 \|u\|^2,$$

so

$$\begin{aligned}\|\nabla f(x)\|^2 &\leq 2\|\nabla f_\mu(x)\|^2 + \frac{2}{\mu^2\kappa} \cdot \left(\frac{1}{4}L_1^2(f)\mu^4\right) \int_{\mathbb{R}^n} \|u\|^6 e^{-\frac{1}{2}\|u\|^2} du \\ &= 2\|\nabla f_\mu(x)\|^2 + \frac{\mu^2}{2}L_1^2(f)M_6.\end{aligned}$$

Finally, from the moment bound (Equation (13)) with $p = 6$, we have:

$$M_6 \leq (n+6)^3,$$

which leads to the desired result:

$$\|\nabla f(x)\|^2 \leq 2\|\nabla f_\mu(x)\|^2 + \frac{\mu^2}{2}L_1^2(f)(n+6)^3.$$

■

B.2 Random Gradient-Free Oracles

Let the random vector $u \in \mathbb{R}^n$ follow a Gaussian distribution with correlation operator B^{-1} . We define the following random gradient-free oracle:

- Sample a random $u \in \mathbb{R}^n$ and return $g_\mu(x) = \frac{f(x + \mu u) - f(x)}{\mu} \cdot Bu$ (19)

We now derive some useful upper bounds. First, for a function f that is differentiable at a point x , we observe that

$$\|g_0(x)\|^2 = \langle \nabla f(x), u \rangle^2 \cdot \|u\|^2 \leq \|\nabla f(x)\|^2 \cdot \|u\|^4$$

Hence, $\mathbb{E}_u(\|g_0(x)\|^2) \stackrel{(13)}{\leq} (n+4)^2 \|\nabla f(x)\|^2$. This bound, however, can be significantly improved.

Theorem B.2. *If f is differentiable at x , then*

$$\mathbb{E}_u(\|g_0(x)\|^2) \leq (n+4)\|\nabla f(x)\|^2 \quad (20)$$

Proof. Let us fix any $\tau \in (0, 1)$. Then, we have:

$$\begin{aligned} \mathbb{E}_u(\|g_0(x)\|^2) &\stackrel{(19)}{=} \frac{1}{\kappa} \int_{\mathbb{R}^n} \|u\|^2 e^{-\frac{1}{2}\|u\|^2} f'(x, u)^2 du \\ &= \frac{1}{\kappa} \int_{\mathbb{R}^n} \|u\|^2 e^{-\frac{\tau}{2}\|u\|^2} f'(x, u)^2 e^{-\frac{1-\tau}{2}\|u\|^2} du \\ &\stackrel{(14)}{\leq} \frac{2}{\kappa \tau e} \int_{\mathbb{R}^n} f'(x, u)^2 e^{-\frac{1-\tau}{2}\|u\|^2} du \\ &= \frac{2}{\kappa \tau (1-\tau)^{\frac{n+1}{2}} e} \int_{\mathbb{R}^n} f'(x, u)^2 e^{-\frac{1}{2}\|u\|^2} du \end{aligned}$$

The minimum of the right-hand side with respect to τ is attained at $\tau_* = \frac{2}{n+4}$. For this value,

$$\tau_*(1-\tau_*)^{\frac{n+2}{2}} = \frac{2}{n+4} \left(\frac{n+2}{n+4} \right)^{\frac{n+2}{2}} > \frac{2}{(n+4)e}$$

Therefore, we obtain the bound:

$$\mathbb{E}_u(\|g_0(x)\|^2) \leq \frac{n+4}{\kappa} \int_{\mathbb{R}^n} f'(x, u)^2 e^{-\frac{1}{2}\|u\|^2} du$$

If f is differentiable at x , then $f'(x, u) = \langle \nabla f(x), u \rangle$, and from equation (9), the result (20) follows immediately. ■

Let us now derive a similar bound for the oracle g_μ .

Theorem B.3. *Let function f be convex. If $f \in C^{(1,1)}$, then*

$$\mathbb{E}_u(\|g_\mu(x)\|^2) \leq \frac{\mu^2}{2} L_1^2(f) (n+6)^3 + 2(n+4)\|\nabla f(x)\|^2 \quad (21)$$

Proof. Let $f \in C^{(1,1)}$. Since

$$\begin{aligned} [f(x + \mu u) - f(x)]^2 &= [f(x + \mu u) - f(x) - \mu \langle \nabla f(x), u \rangle + \mu \langle \nabla f(x), u \rangle]^2 \\ &\stackrel{(3)}{\leq} 2 \left(\frac{\mu^2}{2} L_1(f) \|u\|^2 \right)^2 + 2\mu^2 \langle \nabla f(x), u \rangle^2 \end{aligned}$$

we obtain:

$$\begin{aligned} \mathbb{E}_u(\|g_\mu(x)\|^2) &\leq \frac{\mu^2}{2} L_1^2(f) \cdot \mathbb{E}_u(\|u\|^6) + 2 \cdot \mathbb{E}_u(\|g_0(x)\|^2) \\ &\stackrel{(13)}{\leq} \frac{\mu^2}{2} L_1^2(f) M_6 + 2(n+4) \|\nabla f(x)\|^2 \\ &\stackrel{(20)}{\leq} \frac{\mu^2}{2} L_1^2(f) (n+6)^3 + 2(n+4) \|\nabla f(x)\|^2 \end{aligned}$$

■

Sometimes, it is more convenient to express the right-hand side of inequality (21) in terms of the gradient of the Gaussian approximation.

Lemma B.4. *Let $f \in C^{(1,1)}$. Then, for any $x \in \mathbb{R}^n$ we have*

$$\mathbb{E}_u(\|g_\mu(x)\|^2) \leq 4(n+4) \|\nabla f_\mu(x)\|^2 + 3\mu^2 L_1^2(f) (n+4)^3 \quad (22)$$

Proof. Indeed,

$$\begin{aligned} (f(x + \mu u) - f(x))^2 &= (f(x + \mu u) - f_\mu(x + \mu u) - f(x) \\ &\quad + f_\mu(x) + f_\mu(x + \mu u) - f_\mu(x))^2 \\ &\leq 2(f(x + \mu u) - f_\mu(x + \mu u) - f(x) + f_\mu(x))^2 \\ &\quad + 2(f_\mu(x + \mu u) - f_\mu(x))^2 \end{aligned}$$

Note that

$$\begin{aligned}
|f(x + \mu u) - f_\mu(x + \mu u) - f(x) + f_\mu(x)| &= |f_\mu(x + \mu u) - f(x + \mu u)| + |f_\mu(x) - f(x)| \\
&\stackrel{(15)}{\leq} \frac{\mu^2}{2} L_1(f) \cdot n + \frac{\mu^2}{2} L_1(f) \cdot n \\
&= \mu^2 L_1(f) \cdot n
\end{aligned}$$

and

$$\begin{aligned}
(f_\mu(x + \mu u) - f_\mu(x))^2 &= (f_\mu(x + \mu u) - f_\mu(x) - \mu \langle \nabla f_\mu(x), u \rangle + \mu \langle \nabla f_\mu(x), u \rangle)^2 \\
&\leq 2(f_\mu(x + \mu u) - f_\mu(x) - \mu \langle \nabla f_\mu(x), u \rangle)^2 \\
&\quad + 2\mu^2 \langle \nabla f_\mu(x), u \rangle^2 \\
&\leq 2 \left(\frac{\mu^2}{2} L_1(f) \|u\|^2 \right)^2 + 2\mu^2 \langle \nabla f_\mu(x), u \rangle^2 \\
&= \frac{\mu^4}{2} L_1^2(f) \|u\|^4 + 2\mu^2 \langle \nabla f_\mu(x), u \rangle^2
\end{aligned}$$

Applying (20) to the function f_μ , we obtain $\mathbb{E}_u(\langle \nabla f_\mu(x), u \rangle^2 \|u\|^2) \leq (n+4) \|\nabla f_\mu(x)\|^2$.

Hence,

$$\begin{aligned}
\mathbb{E}_u(\|g_\mu(x)\|^2) &\leq \frac{1}{\mu^2} \mathbb{E}_u((f(x + \mu u) - f(x))^2 \|u\|^2) \\
&\leq 2\mu^2 L_1^2(f) n^2 M_2 + \mu^2 L_1^2(f) M_6 + 4(n+4) \|\nabla f_\mu(x)\|^2 \\
&\leq \mu^2 L_1^2(f) (2n^3 + (n+6)^3) + 4(n+4) \|\nabla f_\mu(x)\|^2 \\
&\leq 3\mu^2 L_1^2(f) (n+4)^3 + 4(n+4) \|\nabla f_\mu(x)\|^2
\end{aligned}$$

It remains to observe that $2n^3 + (n+6)^3 \leq 3(n+4)^3$. ■

B.3 Gaussian Smoothing Gradient Method

We consider the unconstrained convex optimization problem $\min_{x \in \mathbb{R}^n} f(x)$, where x_0 is the initial guess, h_k denotes the step size, and N represents the number of iterations.

Algorithm 1 Gaussian Smoothing Gradient Method

```

1: Input:  $x_0, \mu, h_k, N$ 
2: for  $k = 0$  to  $N$  do
3:   Sample a random vector  $u$ 
4:   Compute  $g_\mu(x_k)$  using (19)
5:   Update:  $x_{k+1} = x_k - h_k g_\mu(x_k)$ 
6: end for
7: return  $x_N$ 

```

The following theorem describes the convergence properties of Algorithm 1 when applied to the problem $\min_{x \in \mathbb{R}^n} f(x)$.

Theorem B.4. *Let $\{x_k\}_{k \geq 0}$ be the sequence generated by Algorithm 1 (Gaussian Smoothing Gradient Method), where $f \in C^{(1,1)}$ satisfies the Polyak–Łojasiewicz (PL) condition with parameters $\alpha > 0$ and $\beta > 0$. Suppose the smoothing parameter $\mu > 0$, and define $g_\mu(x)$ as in (19). Let x_* be a global minimizer of f .*

Then, for any step size $h_k \leq \frac{1}{(\alpha+\beta)(n+4)}$ and scalar $a > 0$, the squared distance to the optimum satisfies:

$$\begin{aligned} \|x_{k+1} - x_*\|^2 &\leq (1 + c_1 h_k) \|x_k - x_*\|^2 \\ &\quad + \left[\frac{h_k^2 \mu^2}{2} L_1^2(f)(n+6)^2 + \frac{h_k \mu^2}{4a} L_1^2(f)(n+3)^3 \right], \end{aligned} \quad (23)$$

where $c_1 = \frac{\alpha\beta}{\alpha+\beta}$.

This bound characterizes the trade-off between optimization progress and the approximation error introduced by zeroth-order smoothing.

Proof. Let $x_{k+1} = x_k - h_k g_\mu(x_k)$ denote the update rule in Algorithm 1. Then,

$$\begin{aligned}\|x_{k+1} - x_*\|^2 &= \|x_k - x_* - h_k g_\mu(x_k)\|^2 \\ &= \|x_k - x_*\|^2 - 2h_k \langle x_k - x_*, g_\mu(x_k) \rangle + h_k^2 \|g_\mu(x_k)\|^2.\end{aligned}$$

Since f_μ is the smoothed approximation and $g_\mu(x_k)$ is an unbiased estimator, we have:

$$\mathbb{E}[g_\mu(x_k)] = \nabla f_\mu(x_k)$$

and therefore:

$$\begin{aligned}\mathbb{E}\langle x_k - x_*, g_\mu(x_k) \rangle &= \langle x_k - x_*, \mathbb{E}[g_\mu(x_k)] \rangle \\ &= \langle x_k - x_*, \nabla f(x_k) \rangle + \langle x_k - x_*, \nabla f_\mu(x_k) - \nabla f(x_k) \rangle \\ &\geq \frac{\alpha\beta}{\alpha + \beta} \|x_k - x_*\|^2 + \frac{1}{\alpha + \beta} \|\nabla f(x_k)\|^2 + \langle x_k - x_*, \nabla f_\mu(x_k) - \nabla f(x_k) \rangle.\end{aligned}$$

Letting $c_1 := \frac{\alpha\beta}{\alpha + \beta}$, the update becomes:

$$\begin{aligned}\|x_{k+1} - x_*\|^2 &\leq (1 - c_1 h_k) \|x_k - x_*\|^2 - \frac{2h_k}{\alpha + \beta} \|\nabla f(x_k)\|^2 + h_k^2 \|g_\mu(x_k)\|^2 \\ &\quad - 2h_k \langle x_k - x_*, \nabla f_\mu(x_k) - \nabla f(x_k) \rangle.\end{aligned}$$

Using the inequality:

$$-2h_k \langle x_k - x_*, \nabla f_\mu(x_k) - \nabla f(x_k) \rangle \leq h_k \left(a \|x_k - x_*\|^2 + \frac{1}{a} \|\nabla f_\mu(x_k) - \nabla f(x_k)\|^2 \right),$$

and applying bounds from (17) and (21), we obtain:

$$\begin{aligned}\|x_{k+1} - x_*\|^2 &\leq (1 - c_1 h_k) \|x_k - x_*\|^2 - \left(\frac{2h_k}{\alpha + \beta} - 2h_k(n + 4) \right) \|\nabla f(x_k)\|^2 \\ &\quad + h_k^2 \cdot \frac{\mu^2}{2} L_1^2(f)(n + 6)^3 + \frac{h_k}{a} \cdot \frac{\mu^2}{4} L_1^2(f)(n + 3)^3 + h_k a \|x_k - x_*\|^2.\end{aligned}$$

Now setting $a = \frac{c_1}{2}$ and choosing step size $h_k \leq \frac{1}{(\alpha+\beta)(n+4)}$, we conclude:

$$\|x_{k+1} - x_*\|^2 \leq (1 + c_1 h_k) \|x_k - x_*\|^2 + \left[\frac{h_k^2 \mu^2}{2} L_1^2(f) (n+6)^2 + \frac{h_k \mu^2}{4a} L_1^2(f) (n+3)^3 \right].$$

■

From the result of the previous theorem, we obtain the recursive inequality (23).

Let us define the following quantities:

$$\begin{aligned} A_k &:= \|x_k - x_*\|^2, \\ D &:= \frac{h_k \mu^2}{2} L_1^2(f) \left[h_k (n+6)^2 + \frac{1}{2a} (n+3)^3 \right], \end{aligned}$$

so that inequality (23) can be rewritten as

$$A_{k+1} \leq (1 + c_1 h_k) A_k + D, \quad \text{for all } k \geq 0. \quad (24)$$

We now expand this recurrence relation iteratively:

$$\begin{aligned} A_1 &\leq (1 + c_1 h_k) A_0 + D, \\ A_2 &\leq (1 + c_1 h_k) A_1 + D = (1 + c_1 h_k)^2 A_0 + (1 + c_1 h_k) D + D, \\ A_3 &\leq (1 + c_1 h_k) A_2 + D = (1 + c_1 h_k)^3 A_0 + (1 + c_1 h_k)^2 D + (1 + c_1 h_k) D + D, \\ &\vdots \\ A_t &\leq (1 + c_1 h_k)^t A_0 + D \sum_{j=0}^{t-1} (1 + c_1 h_k)^j. \end{aligned}$$

Applying the geometric series bound:

$$\sum_{j=0}^{t-1} (1 + c_1 h_k)^j \leq \frac{(1 + c_1 h_k)^t - 1}{c_1 h_k} \leq \frac{(1 + c_1 h_k)^t}{c_1 h_k},$$

we obtain the final inequality:

$$\begin{aligned}
A_t &\leq (1 + c_1 h_k)^t A_0 + \frac{D}{c_1 h_k} \\
&= (1 + c_1 h_k)^t A_0 + \frac{h_k \mu^2}{2c_1 h_k} L_1^2(f) \left[h_k(n+6)^2 + \frac{1}{2a}(n+3)^3 \right] \\
&= (1 + c_1 h_k)^t A_0 + \frac{\mu^2}{2c_1} L_1^2(f) \left[h_k(n+6)^2 + \frac{1}{2a}(n+3)^3 \right].
\end{aligned} \tag{25}$$

This result shows that the optimization error consists of two components: a geometrically decaying term $(1 + c_1 h_k)^t A_0$, and a residual bias term introduced by the Gaussian smoothing approximation, which scales as $\mathcal{O}(\mu^2)$.

C Riemannian Optimization for Zeroth-Order Method

We consider the following Riemannian optimization problem:

$$\min f(x) + r(x), \quad \text{s.t. } x \in \mathcal{M} \tag{26}$$

where \mathcal{M} is a Riemannian submanifold embedded in \mathbb{R}^n , $f : \mathcal{M} \rightarrow \mathbb{R}$ is a smooth and potentially nonconvex function, and $r : \mathbb{R}^n \rightarrow \mathbb{R}$ is a convex and nonsmooth function.

Definition C.1 (Riemannian Gradient). Let f be a smooth function defined on \mathcal{M} . The Riemannian gradient $\text{grad}f(x)$ is a vector in the tangent space $T_x \mathcal{M}$ that satisfies:

$$\left. \frac{d(f(\gamma(t)))}{dt} \right|_{t=0} = \langle v, \text{grad}f(x) \rangle \tag{27}$$

for any $v \in T_x \mathcal{M}$, where $\gamma(t)$ is a smooth curve on \mathcal{M} such that:

$$T_x \mathcal{M} = \{ \gamma'(0) : \gamma(0) = x, \gamma([- \delta, \delta]) \subset \mathcal{M} \text{ for some } \delta > 0, \gamma \text{ is differentiable} \}.$$

Recall that in Euclidean space, a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is said to be L -smooth if

for all $x, y \in \mathbb{R}^n$ it holds that:

$$|f(y) - f(x) - \langle \nabla f(x), y - x \rangle| \leq \frac{L}{2} \|x - y\|^2. \quad (28)$$

We now introduce the Riemannian analogue of L -smoothness. For that, we first define a retraction at $x \in \mathcal{M}$.

Definition C.2 (Retraction). A retraction mapping R_x is a smooth map from the tangent space $T_x \mathcal{M}$ to the manifold \mathcal{M} satisfying:

$$R_x(0) = x, \quad (29)$$

where 0 denotes the zero vector in $T_x \mathcal{M}$, and the differential of R_x at 0 is the identity, i.e.,

$$\left. \frac{dR_x(t\eta)}{dt} \right|_{t=0} = \alpha, \quad \forall \alpha \in T_x \mathcal{M}. \quad (30)$$

In particular, the exponential map Exp_x is a retraction that generates geodesics.

Assumption 1 (L -retraction-smoothness). *There exists a constant $L_g \geq 0$ such that for the function f in problem (26), the following inequality holds:*

$$|f(R_x(\alpha)) - f(x) - \langle \text{grad} f(x), \alpha \rangle| \leq \frac{L_g}{2} \|\alpha\|^2, \quad \forall x \in \mathcal{M}, \alpha \in T_x \mathcal{M}. \quad (31)$$

Definition C.3 (Zeroth-Order Riemannian Gradient). Let $u = Pu_0 \in T_x \mathcal{M}$, where $u_0 \sim \mathcal{N}(0, I_n)$ in \mathbb{R}^n , and $P \in \mathbb{R}^{n \times n}$ is the orthogonal projection matrix onto $T_x \mathcal{M}$. Then u follows a normal distribution $\mathcal{N}(0, PP^T)$ restricted to the tangent space. The zeroth-order Riemannian gradient estimator is defined as:

$$g_\mu(x) = \frac{f(R_x(\mu u)) - f(x)}{\mu} u = \frac{f(R_x(\mu Pu_0)) - f(x)}{\mu} Pu_0. \quad (32)$$

We now consider the smooth optimization problem with $r \equiv 0$, where the objective function f satisfies Assumption 1. We introduce Z0-RGD, the Zeroth-Order

Riemannian Gradient Descent method, and analyze its theoretical complexity. The algorithm is formally described in Algorithm 2.

Algorithm 2 Zeroth-Order Riemannian Gradient Descent(ZO-RGD).

- 1: **Input:** Initial point $x_0 \in \mathcal{M}$, smoothing parameter μ , step size α , fixed number of iteration N .
 - 2: **for** $k = 0$ **to** $N - 1$ **do**
 - 3: Sample a standard Gaussian random vector u_k in $T_{x_k}\mathcal{M}$ via projection.
 - 4: Compute the zeroth-order gradient $g_\mu(x_k)$ by (32)
 - 5: Update $x_{k+1} = R_{x_k}(-\alpha g_\mu(x_k))$.
 - 6: **end for**
-

To analyze the behavior of ZO-RGD, we employ a key geometric tool: a geodesic triangle formed by the current iterate x_k , the next iterate x_{k+1} , and an optimal point x_* . This triangle is illustrated in Figure 6, where each edge corresponds to a geodesic segment and the interior angle A plays an essential role in the analysis of convergence.

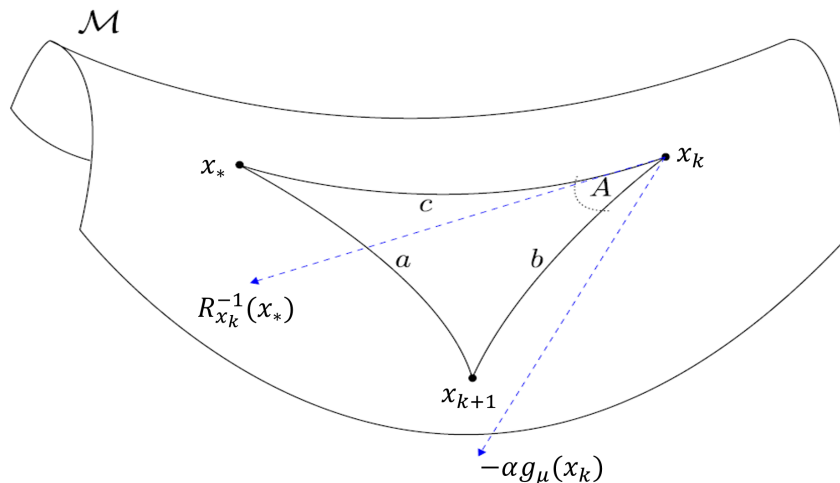


Figure 6: Illustration of the geodesic triangle formed by x_* , x_k , and x_{k+1} .

Let $x_{k+1} := R_{x_k}(-\alpha g_\mu(x_k))$, and consider the geodesic triangle depicted in Figure 6 with x_* , x_k , and x_{k+1} , and sides $a := \text{dist}(x_{k+1}, x_*)$, $b := \text{dist}(x_k, x_{k+1})$, and $c := \text{dist}(x_k, x_*)$. In addition, we have that $bc \cos(A) = \langle -\alpha g_\mu(x_k), R_{x_k}^{-1}(x_*) \rangle$.

If the points form a geodesic triangle in an Alexandrov space with sectional curvature bounded below by κ , then the generalized Law of Cosines gives the inequality:

$$a^2 \leq \frac{c\sqrt{|\kappa|}}{\tanh(c\sqrt{|\kappa|})}b^2 + c^2 - 2bc \cos(A).$$

Applying this to our setting yields:

$$\text{dist}(x_{k+1}, x_*)^2 \leq \text{dist}(x_k, x_*)^2 + 2\alpha \langle g_\mu(x_k), R_{x_k}^{-1}(x_*) \rangle + \zeta \alpha^2 \|g_\mu(x_k)\|^2, \quad (33)$$

where $\zeta := \frac{c\sqrt{|\kappa|}}{\tanh(c\sqrt{|\kappa|})}$ is the curvature-dependent coefficient.

Theorem C.4. *Suppose that $f : \mathcal{M} \rightarrow \mathbb{R}$ is geodesically σ -strongly convex and L -smooth with respect to a retraction R , and that the manifold \mathcal{M} has curvature lower bounded by κ . Let $\{x_k\}_{k=0}^N$ be the sequence generated by Algorithm 2 with fixed step size $\alpha > 0$, smoothing parameter $\mu > 0$, and let $\zeta = \frac{c\sqrt{|\kappa|}}{\tanh(c\sqrt{|\kappa|})}$ where $c = \text{dist}(x_k, x_*)$.*

Then, for all $k \geq 0$, the iterates satisfy the following recursive inequality:

$$\mathbb{E}[\text{dist}(x_{k+1}, x_*)^2] \leq \left(1 + L^2 \alpha^2 \zeta - \alpha \sigma + \frac{\omega \alpha}{2}\right) \text{dist}(x_k, x_*)^2 + \frac{\alpha L^2 \mu^2 (d+3)^3}{2\omega}, \quad (34)$$

where $\omega > 0$ is a tunable constant, and d is the dimension of the ambient space.

Proof. From inequality (33) and the fact that $\text{dist}(x_k, x_{k+1}) = \alpha \|g_\mu(x_k)\|$, we have:

$$\begin{aligned} \text{dist}(x_{k+1}, x_*)^2 &\leq \text{dist}(x_k, x_*)^2 + \zeta \text{dist}(x_k, x_{k+1})^2 - 2\langle -\alpha g_\mu(x_k), R_{x_k}^{-1}(x_*) \rangle \\ &\leq \text{dist}(x_k, x_*)^2 + \zeta \text{dist}(x_k, x_{k+1})^2 + 2\alpha \langle g_\mu(x_k), R_{x_k}^{-1}(x_*) \rangle \end{aligned} \quad (35)$$

By L -smoothness of f , and noting that $\text{grad}f(x_*) = 0$, we obtain:

$$\begin{aligned} \zeta \text{dist}(x_k, x_{k+1})^2 &= \zeta \|x_k - x_{k+1}\|^2 = \zeta \|\alpha \text{grad}f(x_k)\|^2 \\ &= \alpha^2 \zeta \|\text{grad}f(x_k) - \text{grad}f(x_*)\|^2 \leq L^2 \alpha^2 \zeta \|x_k - x_*\|^2 \\ &= L^2 \alpha^2 \zeta \text{dist}(x_k, x_*)^2 \end{aligned}$$

Thus, inequality (35) becomes:

$$\text{dist}(x_{k+1}, x_*)^2 \leq (1 + L^2 \alpha^2 \zeta) \text{dist}(x_k, x_*)^2 + 2\alpha \langle g_\mu(x_k), R_{x_k}^{-1}(x_*) \rangle. \quad (36)$$

Taking expectation conditioned on x_k , we get:

$$\mathbb{E}[\text{dist}(x_{k+1}, x_*)^2] \leq (1 + L^2 \alpha^2 \zeta) \text{dist}(x_k, x_*)^2 + 2\alpha \langle \mathbb{E}[g_\mu(x_k)], R_{x_k}^{-1}(x_*) \rangle. \quad (37)$$

Since f is geodesically strongly convex, it satisfies:

$$-\langle \text{grad} f(x_k), R_{x_k}^{-1}(x_*) \rangle \geq \frac{\sigma}{2} \text{dist}(x_k, x_*)^2, \forall x \in \mathcal{M} \quad (38)$$

By decomposing the gradient estimator error:

$$\begin{aligned} \langle \mathbb{E}[g_\mu(x_k)], R_{x_k}^{-1}(x_*) \rangle &= \langle \text{grad} f(x_k), R_{x_k}^{-1}(x_*) \rangle + \langle \mathbb{E}[g_\mu(x_k)] - \text{grad} f(x_k), R_{x_k}^{-1}(x_*) \rangle \\ &\leq -\frac{\sigma}{2} \text{dist}(x_k, x_*)^2 + \|\mathbb{E}[g_\mu(x_k)] - \text{grad} f(x_k)\| \cdot \text{dist}(x_k, x_*). \end{aligned}$$

From the zeroth-order approximation bound,

$$\|\mathbb{E}[g_\mu(x_k)] - \text{grad} f(x_k)\| \leq \frac{L\mu}{2} (d+3)^{3/2}.$$

Therefore,

$$\mathbb{E}[\text{dist}(x_{k+1}, x_*)^2] \leq (1 + L^2 \alpha^2 \zeta - \alpha \sigma) \text{dist}(x_k, x_*)^2 + \alpha L \mu (d+3)^{3/2} \text{dist}(x_k, x_*). \quad (39)$$

Using the inequality $2xy \leq \omega x^2 + \frac{y^2}{\omega}$ with $x = \sqrt{\alpha} \text{dist}(x_k, x_*)$, $y = \sqrt{\alpha} L \mu (d+3)^{3/2}$, we obtain:

$$\alpha L \mu (d+3)^{3/2} \text{dist}(x_k, x_*) \leq \frac{\omega}{2} \alpha \text{dist}(x_k, x_*)^2 + \frac{\alpha L^2 \mu^2 (d+3)^3}{2\omega}.$$

Substituting into (39), we conclude:

$$\mathbb{E}[\text{dist}(x_{k+1}, x_*)^2] \leq \left(1 + L^2\alpha^2\zeta - \alpha\sigma + \frac{\omega\alpha}{2}\right) \text{dist}(x_k, x_*)^2 + \frac{\alpha L^2\mu^2(d+3)^3}{2\omega}. \quad (40)$$

■

To analyze the long-term behavior of Algorithm 2, define:

$$A_k := \mathbb{E}[\text{dist}(x_k, x_*)^2], \quad D := \frac{\alpha L^2\mu^2(d+3)^3}{2\omega}, \quad c_2 := \sigma - L^2\alpha\zeta - \frac{\omega}{2}.$$

From (40), we obtain the recurrence:

$$A_{k+1} \leq (1 - c_2\alpha)A_k + D. \quad (41)$$

Unfolding this recurrence yields:

$$\begin{aligned} A_1 &\leq (1 - c_2\alpha)A_0 + D, \\ A_2 &\leq (1 - c_2\alpha)^2A_0 + (1 - c_2\alpha)D + D, \\ &\vdots \\ A_t &\leq (1 - c_2\alpha)^tA_0 + D \sum_{j=0}^{t-1} (1 - c_2\alpha)^j. \end{aligned}$$

Using the geometric series bound:

$$\sum_{j=0}^{t-1} (1 - c_2\alpha)^j \leq \frac{1}{c_2\alpha},$$

we obtain:

$$A_t \leq (1 - c_2\alpha)^tA_0 + \frac{D}{c_2\alpha} = (1 - c_2\alpha)^tA_0 + \frac{L^2\mu^2(d+3)^3}{2\omega c_2}. \quad (42)$$

This result establishes that the expected squared Riemannian distance to the

global optimum decays geometrically up to an $\mathcal{O}(\mu^2)$ bias induced by the smoothing approximation.

D Numerical Examples

In this section, we illustrate the convergence properties of the proposed Zeroth-Order Gaussian Smoothing (ZO-GS) and Zeroth-Order Riemannian Gradient Descent (ZO-RGD) methods. We consider both the Euclidean and Riemannian settings.

Experiment 1 (Euclidean Case): Gaussian Smoothing on Least Squares

We consider the unconstrained convex optimization problem

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} \|Ax - b\|^2,$$

where $A \in \mathbb{R}^{m \times n}$ with $m = 100$, $n = 500$, and $b = A\bar{x} + \omega$. Here, $\bar{x} \sim \mathcal{N}(0, I_n)$ and $\omega \sim \mathcal{N}(0, 0.01 \cdot I_m)$. We set the smoothing parameter $\mu = 10^{-4}$, and perform $N = 200,000$ iterations.

To evaluate convergence behavior, we compare three step sizes: $h_k \in \{10^{-8}, 10^{-7}, 10^{-6}\}$. For each step size, the ZO-GS algorithm is repeated 20 times. The empirical mean of the objective value $f(x_k)$ is plotted at logarithmically spaced intervals.

As predicted by theory, a larger step size leads to faster initial convergence, but may result in higher variance and bias in the final solution due to the noise introduced by smoothing and stochastic gradient estimation.

Experiment 2 (Riemannian Case): ZO-RGD on the Sphere

We consider the problem of minimizing a geodesically strongly convex function over the unit sphere $\mathbb{S}^{n-1} \subset \mathbb{R}^n$, given by

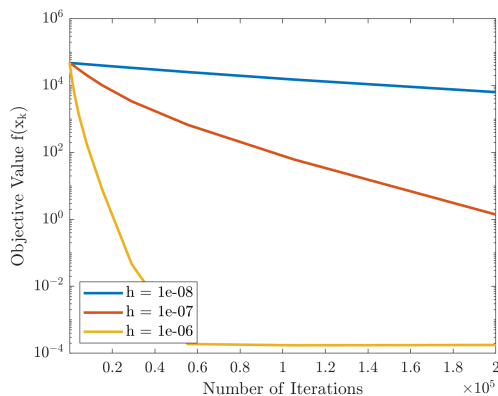
$$\min_{x \in \mathbb{S}^{n-1}} f(x) = -x^\top v,$$

where $v \in \mathbb{R}^n$ is a fixed unit-norm vector. The optimal solution is clearly $x_* = v$, with minimal value $f(x_*) = -1$.

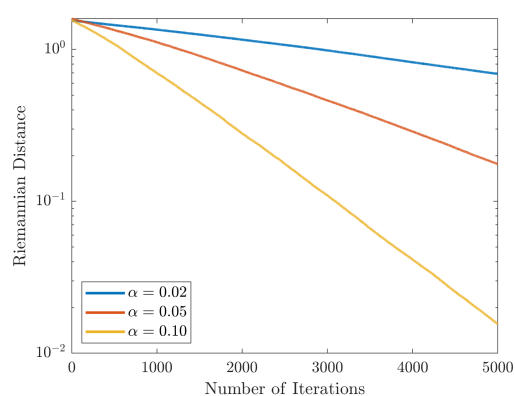
We apply the ZO-RGD (Zeroth-Order Riemannian Gradient Descent) algorithm with fixed step sizes $\alpha \in \{0.02, 0.05, 0.1\}$, a smoothing parameter $\mu = 0.01$, and a total of $N = 5000$ iterations.

For each step size, the algorithm is repeated over 10 independent trials. At each iteration k , we compute the Riemannian distance between the current iterate x_k and the optimal point x_* , which is given by $\text{dist}(x_k, x_*) = \cos^{-1}(x_k^\top x_*)$.

We plot the empirical mean of the Riemannian distance over iterations. As expected, larger step sizes lead to faster initial convergence but may result in a larger steady-state error due to the stochastic nature of gradient estimation and the absence of line search or adaptivity.



(a) Experiment1



(b) Experiment2

Figure 7: The convergence behavior of zeroth-order optimization algorithms in Euclidean and Riemannian settings.

(a) presents the empirical average of objective values $f(x_k)$ obtained from Algorithm 1 with different step sizes h_k in the Euclidean setting. The observed convergence is consistent with the theoretical bound established under the Polyak–Łojasiewicz condition (see Theorem B.4 and Inequality (25)).

(b) shows the empirical average of Riemannian distance $\text{dist}(x_k, x_*)$ over iterations when optimizing over the unit sphere using Algorithm 2. This result aligns with the convergence behavior predicted by the recursive inequality (42), under geodesic strong convexity and retraction-based smoothness (Theorem C.4). Both results highlight the influence of step size on convergence rate and steady-state error in zeroth-order optimization.

E Reference

- [1] Farzin, A. A., and Shames, I.: Minimisation of Polyak-Łojasiewicz Functions Using Random Zeroth-Order Oracles. *arXiv preprint*, arXiv:2405.09106 [math.OC], (2024).
- [2] Gasnikov, A., Dvinskikh, D., Dvurechensky, P., Gorbunov, E., Beznosikov, A., and Lobanov, A.: Randomized Gradient-Free Methods in Convex Optimization. *arXiv preprint*, arXiv:2211.13566 [math.OC], (2022).
- [3] Li, J., Balasubramanian, K., and Ma, S.: Stochastic Zeroth-Order Riemannian Derivative Estimation and Optimization. *arXiv preprint*, arXiv:2003.11238 [math.OC], (2020).
- [4] Maass, A. I., Manzie, C., Nesic, D., Manton, J. H., and Shames, I.: Tracking and Regret Bounds for Online Zeroth-Order Euclidean and Riemannian Optimisation. *arXiv preprint*, arXiv:2010.00211 [math.OC], (2020).
- [5] Li, J., Balasubramanian, K., and Ma, S.: Zeroth-Order Optimization on Riemannian Manifolds. *arXiv preprint*, arXiv:2003.11238 [math.OC], (2020).
- [6] Pougkakiotis, S., and Kalogerias, D. S.: A Zeroth-Order Proximal Stochastic Gradient Method for Weakly Convex Stochastic Optimization. *SIAM Journal on Scientific Computing*, 45(5), A2679–A2702 (2023). Also available as: *arXiv preprint*, arXiv:2205.01633 [math.OC].
- [7] Boumal, N.: *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, (2023).

논문요약

뉴스 데이터와 SVM을 활용한 주식 시장 움직임 예측

손종규

수학과

성균관대학교

본 연구는 조선업종 상장기업을 대상으로 뉴스 감성 분석과 기술적 지표를 결합하여 단기 주가 등락을 예측하는 이진 분류 모델을 구축하였다. KoFinBERT로 뉴스 제목의 감성을 분류하고, 기술적 지표와 함께 SVM 모델에 학습시켰다. 실험 결과, 하락 예측(Class 0) 성능이 상승 예측(Class 1)보다 우수했으며, 피처 중요도 분석에서는 종가, 볼린저 밴드 등 기술 지표가 높은 영향을 보였다.

주제어: 주가 예측, 뉴스 감성 분석, 기술적 지표, SVM, KoFinBERT