

Quantitative Trading Series

**Quantitative and Qualitative Treatments
to Capital Markets**
and related bodies of knowledge

By

HangukQuant

Private Notes,

Quantitative Research

2022~

Quantitative Trading Series

**Quantitative and Qualitative Treatments
to Capital Markets**
and related bodies of knowledge

By

HangukQuant

Private Notes,

Quantitative Research

2022~

DISCLAIMER: the contents of this work are not intended as investment, legal, tax or any other advice, and is for informational purposes only. It is illegal to make unauthorized copies, forward to an unauthorized user or to post this article electronically without express written consent by HangukQuant.

Abstract

This book is designed to be a practical handbook for all finance professionals, practitioner or academic. It is an organization of the various knowledge domains, with a focus on drawing links in the intricate web between the theory and practice of finance that market participants seek to unfold. It contains discussions of trading anomalies, premias and inefficiencies. It contains discussions in discretionary and quantitative trading. Discussion stretches across theoretical work, such as statistical methods, linear algebra and financial mathematics. Applied work in equity research, quantitative research and macroeconomic theory is involved.

This work is attributed to the brilliant writers, academics, scientists and traders before me. Although we have tried to credit the referenced work where relevant, to give a complete reference for its source is impossible. The work has been organized and compiled from various texts, lecture notes, journals, blogs, personal communications and even scraps of scribbled notes from the author's time in college. These contain notes from blogs referencing journals, journals referencing blogs, blogs referencing blogs referring journals - you name it. We apologise if we have failed to credit your work. Other work is original. Readers may reach us at hangukquant@gmail.com. The updated notes are released at hangukquant.substack.com.

Faith is to have believe without seeing. This work is dedicated to those who placed their faith in me. To Jeong(s), Choi, Julian and my dearest friends who have shaped my world view and colored it rainbow.

Keywords:

- Linear Algebra
- Calculus Methods
- Computer Methods
- Global Macro Trading
- Quantitative Research
- Statistics & Probability Theory
- Risk Premia and Market Inefficiencies
- Equities Trading and Other Asset Classes

Table of Contents

Title	i
Abstract	ii
1 Introduction	1
1.1 Guidelines for Reviewing Work	1
2 Ordinary Calculus	2
3 Linear Algebra	3
3.1 Computational Methods in the Euclidean Space	3
3.1.1 Linear Systems	3
3.1.1.1 Elementary Row Operations (EROs)	5
3.1.1.2 Row-Echelon Forms	6
3.1.1.3 Gaussian Elimination Methods	7
3.1.1.4 Homogeneous Linear Systems	9
3.1.2 Matrices	10
3.1.2.1 Operations on Matrices	11
3.1.2.2 Invertibility of Matrices	15
3.1.2.3 Elementary Matrices	17
3.1.2.4 Matrix Determinants	20
3.1.3 Euclidean Vector Spaces	29
3.1.3.1 Finite Euclidean Spaces	29
3.1.3.2 Linear Spans	31
3.1.3.3 Subspaces	33
3.1.3.4 Linear Independence	34
3.1.3.5 Bases	35
3.1.3.6 Dimensions	37
3.1.3.7 Transition Matrices	39
3.1.4 Matrix Vector Spaces	44
3.1.4.1 Row, Column Spaces	45
3.1.4.2 Ranks	47
3.1.4.3 Nullspaces	48
3.1.5 Orthogonality	51
3.1.5.1 Orthogonal Basis	53
3.1.5.2 Best Approximations	55
3.1.5.3 Orthogonal Matrices	57
3.1.6 Diagonalization	63
3.1.6.1 Eigenvalues	63
3.1.6.2 Diagonalization	64
3.1.6.3 Orthogonal Diagonalization	68
3.1.6.4 Quadratic Forms and Conic Sections	69
3.1.7 Linear Transformations	76
3.1.7.1 Ranges and Kernels	78
3.2 Abstract Linear Algebra	82
3.2.1 Vector Spaces	82
3.2.2 Finite Dimensional Vector Spaces	86
3.2.2.1 Vector Spans and Linear Independence	86
3.2.2.2 Bases	89
3.2.2.3 Dimensions of a Vector Space	90
3.3 Matrix Differentiation	92
4 Set Theory	93
4.1 Algebra of Sets	93

5	Probability and Statistical Models	94
5.1	Probability Spaces and Probability Measures	94
5.2	Counting and Combinatorics	97
5.3	Random Variables	100
5.3.1	Expectations of Random Variables	104
5.3.2	Riemann and Lebesgue Integrals	104
5.3.3	Convergence of Integrals	108
5.3.4	Computing Expectations	111
5.3.5	Change of Measure	115
5.3.6	Random Variable Moments	121
5.3.7	Random Variable Co-Movements	123
5.4	Moment Generating Functions	124
5.5	Information and Joint Distributions	125
5.5.1	Independence and Joint Densities	127
5.5.2	Conditioning Probabilities, Densities and Expectations	133
5.5.3	Transformation and Combination of Random Variables	142
5.6	Inferential Statistics	144
5.7	Laws and Basics of Probability Concepts	144
5.8	Sampling Methods	145
5.8.1	Simple Random Sampling	145
5.8.2	Confidence Intervals	146
5.9	Estimation of Ratios	147
5.9.0.1	Estimation of Paired Samples	149
5.10	Method of Moments	150
5.10.1	Consistency of Estimators	151
5.11	Monte Carlo Bootstraps	152
5.12	Method of Maximum Likelihoods	152
5.12.1	Confidence Intervals of Maximum Likelihood Estimates	155
5.12.2	Large Sample Theory of Maximum Likelihoods	155
5.12.3	Least Squares and Maximum Likelihoods	156
5.13	Fisher Information	157
5.14	Efficiency of Estimators and Asymptotic Relative Efficiencies	159
5.14.1	Comparing Asymptotic Relative Efficiency	160
5.15	Sufficient Statistics	161
5.16	Expectations	163
5.16.1	Conditional Expectations	163
5.17	Probability Distributions	164
5.17.1	Bernoulli Trials and Binomial Distribution	164
5.17.2	Hypergeometric Random Distribution	165
5.17.3	Exponential Random Variables	165
5.17.4	Poisson Distribution	166
5.17.5	Poisson Process	166
5.17.6	Gaussian/Normal Distribution	168
5.17.6.1	The Multivariate Case	170
5.17.6.2	Other Gaussian Methods	171
5.17.7	Chi-Squared Distribution χ^2	171
5.17.8	Log-Normal Distribution	172
5.17.9	t Distribution	172
5.17.10	F Distribution	173
5.17.11	Gamma Distribution Γ	173
5.17.12	Beta Distribution	173
5.17.13	Logistic Distributions	174
5.17.14	Rayleigh Distribution	175
5.17.15	Mixture Models	175
5.18	Order Statistics	175
5.18.1	Large Sample Theory on Order Statistics	178
5.19	Methods in Robust Estimation	178
6	Hypothesis Testing, Interval Estimation and Other Tests	180
6.1	Theory of Hypothesis Tests	180
6.2	Generalized Likelihood Ratio Tests	182
6.3	Tests on Normality	184
6.3.1	QQ-Plots	184
6.4	Tests on Mean	184
6.4.1	One-Sample T-Test	184
6.4.2	One-Sample Sign Test	185
6.4.2.1	Normal Approximation to Sign Test and Continuity Corrections	186
6.4.3	Interval Estimation of the Median	186
6.4.4	One-Sample (Wilcoxon) Signed Rank Test	187
6.4.4.1	Distribution of the Wilcoxon Signed Rank Test Statistic	188
6.4.4.2	Relating the Wilcoxon Signed Rank Test to the T-Test	189
6.4.5	Hodge-Lehmann Estimates for Median and Tukey's Method for Confidence Intervals	189

6.4.6	Paired-Sample T-Test	191
6.4.7	Paired Sample Signed Rank Test	191
6.4.8	Two Sample T-Tests	191
6.4.9	Two-Sample Mann Whitney (Wilcoxon Rank Sum) Test	193
6.4.10	Distribution of the Wilcoxon Rank Sum Test Statistic	194
6.5	Location Shift Models	196
6.6	Parametric and Nonparametric Analysis of Variance, ANOVA/F-Test, Kruskal Wallis Test (to be reviewed)	197
6.6.1	F Test	197
6.6.1.1	Groups of Different Size	198
6.6.2	Kruskal Wallis Test	199
6.6.2.1	Distribution of the Kruskal Wallis Test Statistic	200
6.6.3	Bonferroni's Method	200
6.6.4	Tukey's Method	200
6.6.5	Two-Way ANOVA Methods	201
6.7	Correlation Tests	202
6.7.1	Parametric Correlation Test	202
6.7.2	Nonparametric Spearman Correlation Test	202
6.8	Goodness of Fit Tests	203
6.8.1	Kolmogorov Smirnov Test	204
7	Statistical Learning	206
7.1	Supervised Learning	206
7.2	Generalized Linear Models	206
7.2.1	Nearest Neighbors	207
7.2.2	Least Squares <i>vs</i> Nearest Neighbours	207
7.2.3	Regression Functions, Classifiers and Prediction Errors	208
7.3	Local Methods in High Dimensions	210
7.4	Bias, Variance and Errors	210
7.5	Statistical Models	211
7.5.1	Supervised Statistical Models	211
7.6	Classes of Restricted Estimators	212
7.6.1	Roughness Penalty and Bayesian Methods	212
7.6.2	Kernel Methods and Local Regression	212
7.6.3	Basis Functions and Dictionary Methods	213
7.7	Model Selection, Bias-Variance Tradeoff	214
7.8	Least Squares Regression Methods	214
7.8.1	Simple Least Squares	214
7.8.1.1	Assumptions of the Simple Linear Equation	215
7.8.1.2	Model Fitting	215
7.8.1.3	Model Properties and Variance of Estimates	216
7.8.1.4	Assumptions of the Analysis of Model on Simple Linear Equations	217
7.8.1.5	Test of Significance on Regression Coefficients	217
7.8.1.6	Test of Significance on Regression Model and ANOVA Methods	217
7.8.1.7	Confidence Intervals on Parameters and Variance Estimates	218
7.8.1.8	Confidence Intervals and Prediction Intervals on Response	219
7.8.1.9	No Intercept Models	219
7.8.1.10	Coefficient of Determination, R^2	220
7.8.1.11	Maximum Likelihood Estimators vs Simple Least Squares	220
7.8.2	Multiple Least Squares	221
7.8.2.1	Interaction Effects	221
7.8.2.2	Assumptions and Model Notations	221
7.8.2.3	Model Fitting	221
7.8.2.4	Model Properties and Variance of Estimates	222
7.8.2.5	Assumptions for Analysis of Multiple Least Squares Regression	223
7.8.2.6	Significance Tests for Regression Coefficients by t-tests and Partial Sum of Squares Method	223
7.8.2.7	Confidence Interval for Regression Coefficient Estimates	224
7.8.2.8	Confidence Interval and Prediction Intervals of Estimates on Mean Response and Response Variables	225
7.8.2.9	Significance Tests for Regression Model	225
7.8.2.10	Coefficients of Determination and Adjustments	226
7.8.2.11	Interpretation of Model and Coefficients	226
7.8.2.12	Regressor Variable Hull and Extrapolation of the Input Space	226
7.8.2.13	Standardization of Model	227
7.8.2.14	Indicator Functions	227
7.8.3	Adequacy of The Least Squares Method	228
7.8.3.1	Residual Analysis	228
7.8.3.1.1	Leverage Values, Influential Values and the Variance of Residuals	228
7.8.3.2	Standardization of Residuals	229

7.8.3.3	Checking Normality Assumptions	229
7.8.3.4	Note on Time Series Data	230
7.8.3.5	Outliers and Influential Data	230
7.8.3.6	Lack of Fit	230
7.8.3.7	Multicollinearity	231
7.8.3.7.1	Detection of multicollinearity - Plots, Variance Inflation Factors and Eigensystem Analysis	232
7.8.4	Correction of Inadequacies	233
7.8.4.1	Transformation of Response-Regressor	233
7.8.4.1.1	Box-Cox Method	233
7.8.4.1.2	Box-Tidwell Method	233
7.8.4.2	Ridge Regression	234
7.8.4.3	Principal Component Regression	234
7.8.5	Weighted Least Squares	234
7.8.6	Generalized Least Squares	235
7.8.7	Variable Selection Methods	236
7.8.7.1	Selection Criterion	237
7.8.7.1.1	Akaike Information Criteria	237
7.8.7.1.2	Bayesian Information Criteria	237
7.8.7.2	Computational Methods - Brute Force and Stepwise Greedy Solutions	237
7.8.7.2.1	Brute Force Method	237
7.8.7.2.2	Forward Selection Method	237
7.8.7.2.3	Backward Elimination Method	238
7.8.7.2.4	Stepwise Regression Method	238
7.9	Rank Methods for Linear Regression	238
7.10	Nonparametric Density Curve Estimation	240
7.10.1	Histogram	240
7.10.2	Naive Kernel Density Estimator	240
7.10.3	Kernel Density Estimator	241
7.11	Kernel Regression	243
7.11.1	Nadaraya–Watson Kernel Regression	243
8	Randomization and Simulation	244
8.1	Permutations	244
9	Utility Theory	246
9.1	Utility Functions	246
9.1.1	CRRA Utility	246
10	Statistical Finance	247
10.1	Simulation and Resampling Methods	247
10.1.1	Destruction of Patterns in Prices and Bars by Permutation	247
10.1.1.1	Permuting Price Data	247
10.1.1.2	Permuting Bar (OHLCV) Data	248
10.1.1.2.1	Cautionary Note on the Walkforward Data Shuffle	252
10.2	Probabilistic Analysis of Trading Systems	252
10.2.1	Parametric and Non-Parametric Methods for Tests of Location	253
10.2.2	Monte Carlo Permutation Methods for Arbitrary Performance Criterion	254
10.2.2.1	p-value in in-sample and overfit detection by data shuffle	255
10.2.2.2	p-value of asset timing in oos by decision shuffle	256
10.2.2.3	p-value of asset picking in oos by decision shuffle	256
10.2.2.4	p-value of trader skill in oos by data shuffle	257
10.2.2.5	p-value of signal families in oos by data shuffle and bias adjustments	257
10.2.2.5.1	p-value of 1-best signal	258
10.2.2.5.2	p-value of k-combined signals	258
10.2.2.5.3	p-value of k-marginal signals by bounding	259
10.2.2.5.4	p-value of k-marginal signals by greedy selection and FER control	259
10.2.3	Monte Carlo Bootstraps	261
11	Portfolio Management	262
11.1	Introduction and Problem Settings	262
11.1.1	Returns	262
11.1.2	Risk and Standard Deviations	263
11.1.3	VaR, Conditional VaR	264
11.1.4	Alternative Measures of Portfolio Risk	265
11.1.5	Risk-Adjusted Returns	266
11.1.6	A Basket of Assets	267
11.1.7	Computations for the Portfolio	268
11.2	Dynamic Signal Weighting	268
11.2.1	Elegant Mathematics, Poor Economics - MPT	269
11.2.2	Risk-Free Rates and the Capital Market Line	272

11.3	Utility functions and Indifference Curves	274
11.4	Factor Modelling	275
11.4.1	Time Series Factor Models	277
11.4.2	Cross Sectional Models	279
11.4.3	Statistical Models	281
11.4.4	Simplified Markowitz Formulations under the Factor Model Analysis	281
11.4.5	Factor Hedging and Attribution	282
11.4.5.1	Single/Multiple Factor Hedging	282
11.5	Extension of the Classical Mean-Variance Framework for Various Constraints and Utilities	284
11.5.1	Practical Constraints and Modifications to Objectives	284
11.5.2	Extension of the Utility Function for Higher Moments of Portfolio Returns	287
11.5.2.1	Polynomial Goal Programming, Fabozzi et al. [4]	288
11.5.3	Numerical Solving of the Extension to N Asset Mean-Variance Portfolios with Long Constraints and Linear Costs	289
11.5.4	Numerical Solving of the Extension to N Multi-Asset Alpha Mean-Variance Portfolios with Long Constraints and Linear Costs	292
11.6	Robust Estimation Methods and Algorithms	294
11.6.1	Robust Estimators for Center	296
11.6.2	Robust Estimators for Spread	296
11.6.3	Robust Estimation of Returns	297
11.6.4	Robust Estimation of High Dimensional Covariance and Correlation Matrices	298
11.6.4.1	Difficulty in Estimation of the Covariance Matrix	299
11.6.4.2	Random Matrix Theory	299
11.6.4.3	Change in Centre of Mass	300
11.6.4.4	Ridge Estimator	300
11.6.4.5	Ledoit Wolf Constant Correlation	300
11.6.4.6	EPO Shrinkage	300
11.6.5	Bayesian Methods	300
12	Stochastic Calculus	301
12.1	Brownian Motion	301
12.1.1	Random Walk	301
12.1.2	Limiting Distributions and Derivation of Log-Normality	302
12.1.3	Brownian Motion	306
12.1.3.1	Brownian Motion Filtration	307
12.1.3.2	Variation of the Brownian Motion	308
12.1.3.3	Volatility of an Exponentiated Brownian Motion	311
12.1.4	Brownian Motion as Markov Process	312
12.1.5	First Passage Time Distribution	314
12.1.6	Reflection Principle	317
12.1.7	Brownian Density	318
12.2	Stochastic Calculus	319
12.2.1	Ito Integral	319
12.2.2	Ito Doebelin Formula	326
12.2.3	Black Scholes Merton	339
12.2.3.1	Evolution of Portfolio Value	339
12.2.3.2	Evolution of Option Value	340
12.2.3.3	Deriving the BSM PDE	341
12.2.4	Verifying the Black Scholes Merton Solution	341
12.2.5	Option Greeks	342
12.2.6	Put-Call Parity	350
12.2.7	Multivariable Stochastic Calculus	351
12.2.8	Levy's Theorem	354
12.2.9	Gaussian Processes	357
12.2.10	Brownian Bridge	365
12.2.10.1	Joint Distributions of the Brownian Bridge	367
12.2.11	Brownian Bridge as Conditioned Brownian Motion	370
12.3	Risk-Neutrality	371
12.3.1	Risk Neutral Measure, Generalized Geometric Brownian Motion	372
12.3.2	Risk Neutral Measure, Value of Portfolio Process	373
12.3.3	Risk Neutral Measure, Derivative Pricing	374
12.3.4	Risk Neutral Measure, Obtaining the Black-Scholes-Merton Form	374
12.3.5	Martingale Representation Theorem	378
12.3.5.1	One Brownian Motion Martingale Representation	378
12.3.5.2	Single Stock Hedging	380
12.3.6	Fundamental Theorem of Asset Pricing	380
12.3.6.1	Multidimensional Market Model	381
12.3.6.2	Existence of Risk-Neutral Measure	383
12.3.6.3	Uniqueness of Risk-Neutral Measures	388
12.3.7	Dividend Paying Stocks	391
12.3.7.1	Continuous Dividends	391
12.3.7.2	European Call Pricing for Continuous Dividends	392

12.3.7.3	Lump Sum Dividends	397
12.3.7.4	European Call Pricing for Lump Sum Dividends	397
12.3.8	Forwards and Futures	398
12.3.8.1	Forwards	398
12.3.8.2	Futures	399
12.3.9	Difference in Valuations Between Forwards and Futures	401
12.4	Partial Differential Equations	402
12.4.1	Stochastic Differential Equations	402
12.4.2	Markov Property	404
12.4.3	Feynman-Kac Theorem	404
12.4.4	Applications to Interest Rate Models	412
12.4.5	Multidimensional Feynman-Kac Theorem	420
12.5	Exotic Options	422
12.5.1	Up-and-Out Call	425
12.5.2	Lookback Options	429
12.5.3	Asian Options	435
12.6	American Derivative Securities	441
12.6.1	Perpetual American Put	442
12.6.2	Finite Expiration American Put	453
12.6.3	Finite Expiration American Call	455
12.7	Change of Numeraire	459
12.7.1	Foreign, Domestic Risk-Neutral Measures	466
12.7.1.1	Domestic Risk-Neutral Measure	467
12.7.1.2	Foreign Risk-Neutral Measure	468
12.7.2	Forward Exchange Rate	472
12.7.3	Garman-Kohlhagen	472
12.7.4	Exchange Rate Put-Call Duality	473
12.7.5	Forward Measure	473
12.7.5.1	Zero-Coupon Bond Numeraire	474
12.7.6	Stochastic Rate Option Pricing	474
12.8	Term Structure Models	479
12.8.1	Affine-Yield Models	480
12.8.1.1	Two-Factor Vasicek Model	480
12.8.1.2	Two-Factor CIR Model	491
12.8.1.3	Mixed Model	493
12.8.2	Heath-Jarrow-Morton	493
12.8.2.1	Dynamics of Forward Rates and Bond Prices	494
12.8.2.2	Heath-Jarrow-Morton under Risk-Neutrality	496
12.8.2.3	Relation to the Affine-Yield Models	497
12.8.2.4	Implementation of HJM	498
12.8.3	Forward LIBOR Model	500
12.8.3.1	(Forward) LIBOR	500
12.8.3.2	Backset LIBOR Contract	500
12.8.3.3	Black Caplet Formula	501
12.8.3.4	Relating Forward LIBOR and ZCB Volatility	502
12.8.3.5	Forward LIBOR Term Structure Model	503
12.9	Calculus for Jump Processes	506
12.9.1	Compensated and Compound Poisson Processes	506
12.9.2	Moment Generating Function for the Compound Poisson Process	508
12.9.3	Integrals and Differentials of Jump Processes	510
12.9.4	Change of Measure for Jump Processes	520
12.9.5	Option Pricing under the Jump Model of $S(t)$ driven by $N(t)$	529
12.9.6	Option Pricing under the Jump Model of $S(t)$ driven by $W(t)$ and $Q(t)$	532
13	Risk Premiums	538
13.1	Formalism	538
13.1.1	Microeconomic Foundations of Risk Premiums	540
13.1.1.1	One-Period Model	540
13.1.2	Idiosyncratic Risk	542
13.1.2.1	Continuous Time Models	542
13.1.3	Interest Rates	544
13.1.4	Expected Returns	545
13.1.4.1	Central Pricing Equation and the Capital Asset Pricing Model	546
14	Historical Notes	547
15	Alternative Data	548
15.0.5	Practice of Working with Alternative Data	548
15.0.6	Sources of Alternative Data and Their Respective Handlers	548
16	Conclusions	549

A Russian Doll Tester	1
B Market Historians	1
B.1 1960	1
B.2 1961	1
B.3 1962	1
B.4 1963	2
B.5 1964	2
B.6 1965	2
B.7 1966	3
B.8 1967	3
B.9 1968	3
B.10 1969	4
B.11 1970	4
B.12 1971	4
B.13 1972	4
B.14 1973	5
B.15 1974	5
B.16 1975	5
B.17 1976	6
B.18 1977	6
B.19 1978	6
B.20 1979	7
B.21 1980	7
B.22 1981	7
B.23 1982	8
B.24 1983	8
B.25 1984	8
B.26 1985	9
B.27 1986	9
B.28 1987	9
B.29 1988	10
B.30 1989	10
B.31 1990	10
B.32 1991	11
B.33 1992	11
B.34 1993	11
B.35 1994	12
B.36 1995	12
B.37 1996	12
B.38 1997	13
B.39 1998	13
B.40 1999	14
B.41 2000	14
B.42 2001	14
B.43 2002	15
B.44 2003	15
B.45 2004	15
B.46 2005	16
B.47 2006	16
B.48 2007	17
B.49 2008	17
B.50 2009	17
B.51 2010	18
B.52 2011	18
B.53 2012	19
B.54 2013	19
B.55 2014	20
B.56 2015	20
B.57 2016	20
B.58 2017	21
B.59 2018	21
B.60 2019	22
B.61 2020	22
B.62 2021	22
B.63 2022	23
B.64 2023	23
C CODE References	1

Chapter 1

Introduction

1.1 Guidelines for Reviewing Work

The following are the stages of alpha formulations.

Idea 1 (This means to further explore the idea creatively. This is a precursor to a **Test**.).

Test 1 (This refers to parameterized research idea that is to be verified as a **Strategy**.).

Strategy 1 (This explores the implementation and characteristics of a **Test**.).

The following are the stages of theoretical formulations.

Definition 1 (Standard conventions and formal nomenclature are introduced.).

Problem 1 (A formalization of the problem statement is provided).

Exercise 1 (An example or working problem to demonstrate concepts discussed).

The following are stages of theoretical derivations

Lemma 1 (An important result used as is or for other derivations.).

Corollary 1 (An important aside of the theoretical work.).

Theorem 1 (A central result with derivations).

Result 1 (A central result without proof.).

The following are for declarative statements.

Proposition 1 (An opinion of sorts.).

Fact 1 (A statement of (almost) undeniable truth.).

Chapter 2

Ordinary Calculus

Theorem 2 (Integration By Parts). *The integration by parts formula takes form*

$$\int u dv = uv - \int v du \tag{1}$$

Theorem 3 (L'Hopital's Rule).

Chapter 3

Linear Algebra

Here we discuss concepts in linear algebra - notably the literature on this subject is divided into two different schools. One introduces linear algebra as the mathematics and computation of multiply defined linear equations. Here the focus is on teaching linear algebra as a tool for manipulation and computation in multi-dimensional spaces. Determinants are introduced early on, and focuses are on matrix operations. The second approach is to treat matrices as abstract objects, laying focus to the structure of linear operators and vector spaces. Determinants and matrices are only introduced later. Here we provide both - the first will focus on the linear algebraic manipulation of matrices on finite-dimensional, Euclidean spaces. The second treatment will focus on the underlying mathematics of the structure of linear operators and their properties, including the mathematics in infinite dimensional vector spaces and over complex fields. Some of these treatments and notes on Linear Algebra herein are adapted from the texts from Ma et al. [8], Axler [2] and Roman [12].

3.1 Computational Methods in the Euclidean Space

3.1.1 Linear Systems

Definition 2 (Linear Equation). *A linear equation is one in which for variables $\{x_1, \dots, x_n\}$, equation takes form*

$$\sum_{i=1}^n a_i x_i = b \tag{2}$$

where $a_i \in \mathbb{R}, i \in [n]$ and $b \in \mathbb{R}$.

Definition 3 (Zero Equation). *A zero equation is a linear equation (see Definition 2) where all $i \in [n], a_i = 0$ and $b = 0$. That is,*

$$0x_1 + 0x_2 + \dots + 0x_n = 0. \tag{3}$$

The variables $x_i, i \in [n]$ in Definition 2 are not known and it is our task to solve for the solutions to these. The number of variables defines the dimensionality of our problem setting. For instance, see that the equation $ax + by + cz = d$ specify variables in the three-dimensional space $(x, y, z) \in \mathbb{R}^3$. For instance, the linear equation $z = 0$ specifies an xy-plane inside the xyz-space.

Definition 4 (Solution and Solution Sets to a Linear Equation). A solution to a linear equation (see Definition 2) is a set of numbers $\{x_1 = s_1, x_2 = s_2, \dots, x_n = s_n\}$ that satisfies the linear equation $s.t.$

$$\sum_{i=1}^n a_i s_i = b. \quad (4)$$

The set of all such solutions is called a solution set to the equation. When the solution set is expressed by equations representing exactly the equations in the solution set, these set of expressions are known as the general solution.

For instance, in the xy -space, solutions to the equation $x + y = 1$ are points taking the form $(x, y) = (1-s, s)$ where $s \in \mathbb{R}$. In the xyz -space, the solutions to the same equation are points $(x, y, z) = (1-s, s, t)$ where $s, t \in \mathbb{R}$. The solution set form points on a plane. The solution set to the zero equation (see Definition 3) is the entire space \mathbb{R}^n corresponding to the number of dimensions in the linear equation. The solution set to $\sum_i^n 0x_i \neq 0$ is \emptyset .

Definition 5 (Linear System). A finite set of m equations in n variables x_1, \dots, x_n is called a linear system and may be represented

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = b_i, \quad i \in [m] \quad (5)$$

where $a_{ij}, i \in [m], j \in [n] \in \mathbb{R}$.

Definition 6 (Zero System). A zero system is a linear system (see Definition 5) where all the constants $a_{ji}, b_j, i \in [n], j \in [m]$ are zero.

Definition 7 (Solution and Solution Sets to a Linear System). A solution to a linear system (see Definition 5) is a set of numbers $\{x_1 = s_1, x_2 = s_2, \dots, x_n = s_n\}$ that satisfies all linear equations (i.e)

$$\sum_{i=1}^n a_{ji}s_i = b_j, \quad j \in [m] \quad (6)$$

The set of all such solutions is called a solution set to the system. When the solution set is expressed by equations representing exactly the equations in the solution set, these set of expressions are known as the general solution.

Definition 8 (Consistency of Systems). A system of linear equations that has solution set $\neq \emptyset$ is said to be consistent. Otherwise it is inconsistent.

Every system of linear equations will either be consistent or inconsistent. Consistent systems have either a unique solution or infinitely many solutions.

Exercise 2. Show that a linear system $Ax = b$ has either no solution, only one solution or infinitely many.

Proof. If the linear system is not consistent then it must have no solution. Otherwise, it may have a unique solution, or more than one solution. Suppose there are two solutions $u \neq v$ and $Au = Av = b$. Then we may write

$$A(tu + (1-t)v) = tAu + (1-t)Av = tb + (1-t)b = tb + b - tb = b. \quad (7)$$

This is valid for all $t \in \mathbb{R}$, and so we have infinitely many solutions. \square

For example, a system of two linear equations in two-dimensional space each representing a line has infinite solutions if they are the same line, no solution if they are parallel but different lines, and exactly one solution otherwise.

Exercise 3. In the xyz -space, the two equations

$$a_1x + b_1y + c_1z = d_1, \quad (E_1) \tag{8}$$

$$a_2x + b_2y + c_2z = d_2, \quad (E_2) \tag{9}$$

where $\exists a_1, b_1, c_1 \neq 0 \wedge \exists a_2, b_2, c_2 \neq 0$ represents two planes. The solution to the system is the intersection between the two planes. Logicize that there is either no solution ($E_1 // E_2$) or infinite number of solutions ($(E_1 = E_2) \vee (E_1 \text{ intersects } E_2 \text{ on a line})$).

3.1.1.1 Elementary Row Operations (EROs)

Definition 9 (Augmented Matrix Representation of Linear Systems). See that the system of linear equations (Definition 5) given

$$\forall j \in m, \quad \sum_{i=1}^n a_{ji}x_i = b_j \tag{10}$$

may be represented by the rectangular array of numbers

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array} \right] \tag{11}$$

and we call this the augmented matrix of the system. We denote this $(A|b)$. Sometimes, we omit this representation and just assign a single letter, say M , to represent the entire matrix.

Definition 10 (Elementary Row Operations). When we solve for a linear system, we implicitly or explicitly perform the following operations; *i*) multiply equation by some non-zero $k \in \mathbb{R}$, *ii*) interchange two equations, *iii*) add a multiple of one equation to another. In the augmented matrix (see Definition 9), these operations correspond to multiplying a row by a non-zero constant, swapping two rows and adding a multiple of one row to another row respectively. These three operations are collectively known as the elementary row operations. We adopt the following notations

1. $kR_i \equiv$ multiply row i by k .
2. $R_i \leftrightarrow R_j \equiv$ swap rows i, j .
3. $R_j + kR_i \equiv$ add k times of row i to row j .

Definition 11 (Row Equivalent Matrices). Two matrices A, B are said to be row equivalent if one may be obtained by another from a series of EROs. We denote this by

$$A \stackrel{\mathcal{R}}{\equiv} B. \tag{12}$$

Theorem 4 (Solution Sets of Row Equivalent Augmented Matrix Represented Linear Systems). Two linear systems (Definition 5) with augmented matrix representations $(A|b), (C|d)$ have the same solution set if $(A|b) \stackrel{\mathcal{R}}{\equiv} (C|d)$.

Proof. See proof in Exercise 14 using block matrix notations. □

3.1.1.2 Row-Echelon Forms

Definition 12 (Leading Entry). *The first non-zero number in a row of the matrix is said to be the leading entry of the row.*

Definition 13 (Zero Row). *Let the row representing a zero equation (see Definition 3) be called the zero row.*

Definition 14 (Zero Column). *Let the column representing all zero coefficients in the representative linear system for some variable (see Definition 6) be called the zero column. That is, the column has all zeros.*

Definition 15 (Row-Echelon Form (REF)). *A matrix is said to be row-echelon if the following properties hold:*

1. *Zero rows (Definition 13) are grouped at the bottom of the matrix.*
2. *If any two successive rows are non-zero rows, then the higher row has a leading entry (Definition 12) occurring at a column that is to the left of the lower row.*

For matrix A , we denote its matrix REF as $REF(A)$.

Definition 16 (Pivot Points/Columns). *The leading entries (Definition 12) of row-echelon matrices (Definition 15) are called pivot points. The column of a row-echelon form containing a pivot point is called a pivot column, and is otherwise a non-pivot column.*

Definition 17 (Reduced Row-Echelon Form (RREF)). *A reduced row-echelon-form matrix is a row-echelon-form matrix that has*

1. *All leading entries of non-zero row equal to one. (Definitions 12 and 13)*
2. *In each pivot column, all entries other than the pivot point is equal to zero. (Definition 16)*

For matrix A , we denote its matrix RREF as $RREF(A)$.

Note that a zero system is an REF (and also an RREF) by the Definitions 15 and 17. We show that obtaining the REF and RREF gives us an easy way to obtain the solution set to a linear system.

Exercise 4 (Finding solutions to REF, RREF Representations of Linear Systems; Back-Substitution Method). *Find the solution set to the linear systems represented by the following augmented matrices. (see Definitions 9, 5 and 4)*

- 1.

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array} \right] \quad (13)$$

- 2.

$$\left[\begin{array}{ccccc|c} 0 & 2 & 2 & 1 & -2 & 2 \\ 0 & 0 & 1 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 & 2 & 4 \end{array} \right] \quad (14)$$

3.

$$\left[\begin{array}{cccc|c} 1 & -1 & 0 & 3 & -2 \\ 0 & 0 & 1 & 2 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right] \quad (15)$$

4.

$$\left[\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right] \quad (16)$$

5.

$$\left[\begin{array}{cc|c} 3 & 1 & 4 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{array} \right] \quad (17)$$

Proof. 1. It is easy to see that $x_1 = 1, x_2 = 2, x_3 = 3$ is the unique solution this linear system.

2. Since this represents the linear system

$$2x_2 + 2x_3 + x_4 - 2x_5 = 2, \quad (18)$$

$$x_3 + x_4 + x_5 = 3, \quad (19)$$

$$2x_5 = 4, \quad (20)$$

solve. We let the solutions to variables of non-pivot columns be arbitrary. That is, $x_1 \in \mathbb{R}$. The third equation says $x_5 = 2$. Substituting into the second equation, get

$$x_3 + x_4 + 2 = 3, \quad (21)$$

so $x_3 = 1 - x_4$. Substituting into first equation,

$$2x_2 + 2(1 - x_4) + x_4 - 2 \cdot 2 = 2, \quad (22)$$

so $x_2 = 2 + \frac{1}{2}x_4$. So there are two free parameters, and we arrive at the general solution $(x_1, x_2, x_3, x_4, x_5) = (s, 2 + \frac{1}{2}t, 1 - t, t, 2)$, where $s, t \in \mathbb{R}$. This technique is known as the back-substitution method.

3. By the same back-substitution method, arrive at the general solution $(x_1, x_2, x_3, x_4) = (-2 + s - 3t, s, 5 - 2t, t)$ where $s, t \in \mathbb{R}$.

4. The solution set is $(r, s, t) = \mathbb{R}^3$.

5. This system is inconsistent! (Definition 8)

□

3.1.1.3 Gaussian Elimination Methods

Let $A \stackrel{\mathcal{R}}{\equiv} R$. If R is (R)REF, R is said to (reduced) row-echelon form of A and A is said to have (R)REF form R .

Theorem 5 (Gaussian Elimination/Row Reduction and Gauss-Jordan Elimination). *We outline the algorithm to reduce a matrix A to its REF B .*

1. Locate the leftmost non-zero column (see Definition 14).
2. If this happens to be the top-most column, then continue. Else, swap the top row with the row corresponding to the leading entry (Definition 12) found in the previous step.
3. For each row below the top row, add a suitable multiple so that all leading entries below the leading entry of the top row equals zero.
4. From the second row onwards, repeat algorithm from step 1 applied to the submatrix until REF is obtained.

To further get a RREF from REF obtained,

5. Multiply a suitable constant to each row so that all the leading entries become one.
6. From the bottom row onwards, add suitable multiples of each row such that all rows above the leading entries at pivot columns (Definition 16) are all zero.

Steps 1 – 4 are known as *Gaussian Elimination*. Obtaining the RREF via Steps 1 – 6 is known as *Gauss-Jordan elimination*.

Exercise 5. Obtain the RREF of the following augmented matrix

$$\left[\begin{array}{ccccc|c} 0 & 0 & 2 & 4 & 2 & 8 \\ 1 & 2 & 4 & 5 & 3 & -9 \\ -2 & -4 & -5 & -4 & 3 & 6 \end{array} \right] \quad (23)$$

via *Gauss-Jordan Elimination* (see Theorem 5).

Proof. Recall the notations for EROs (see Definition 10). We perform the following steps;

$$\left[\begin{array}{ccccc|c} 1 & 2 & 4 & 5 & 3 & -9 \\ 0 & 0 & 2 & 4 & 2 & 8 \\ -2 & -4 & -5 & -4 & 3 & 6 \end{array} \right] \quad R_1 \leftrightarrow R_2, \quad (24)$$

$$\left[\begin{array}{ccccc|c} 1 & 2 & 4 & 5 & 3 & -9 \\ 0 & 0 & 2 & 4 & 2 & 8 \\ 0 & 0 & 3 & 6 & 9 & -12 \end{array} \right] \quad R_3 + 2 \cdot R_1, \quad (25)$$

$$\left[\begin{array}{ccccc|c} 1 & 2 & 4 & 5 & 3 & -9 \\ 0 & 0 & 2 & 4 & 2 & 8 \\ 0 & 0 & 0 & 0 & 6 & -24 \end{array} \right] \quad R_3 - \frac{3}{2} \cdot R_2, \quad (26)$$

$$\left[\begin{array}{ccccc|c} 1 & 2 & 4 & 5 & 3 & -9 \\ 0 & 0 & 1 & 2 & 1 & 4 \\ 0 & 0 & 0 & 0 & 1 & -4 \end{array} \right] \quad \frac{1}{2}R_2, \quad \frac{1}{6}R_3, \quad (27)$$

$$\left[\begin{array}{ccccc|c} 1 & 2 & 4 & 5 & 0 & 3 \\ 0 & 0 & 1 & 2 & 0 & 8 \\ 0 & 0 & 0 & 0 & 1 & -4 \end{array} \right] \quad R_2 - 1 \cdot R_3, \quad R_1 - 3 \cdot R_3, \quad (28)$$

$$\left[\begin{array}{ccccc|c} 1 & 2 & 0 & -3 & 0 & -29 \\ 0 & 0 & 1 & 2 & 0 & 8 \\ 0 & 0 & 0 & 0 & 1 & -4 \end{array} \right] \quad R_1 - 4 \cdot R_2. \quad (29)$$

1

□

¹we thank reader Irena for the correction of errata in the Gaussian Elimination workings.

Result 2 (REF and their Interpretations for Solution Sets). *Consider the REF $(A|b)$ augmented matrix form (see Definition 9). Note that every matrix has a unique RREF but can have many different REFs. If a linear system is not consistent (Definition 8), then the last column of the REF form of the augmented matrix is a pivot column. In other words, there will be a row representing an equation where $\sum_i^n 0x_i = c$, but $c \neq 0$. There is no solution to this linear system. A consistent linear system has a unique solution if except the last column b , every column is a pivot column. This system has as many variables in the linear system as the number of nonzero rows in the REF. If there exists a non-pivot column in the REF that is not the last one (b), then this consistent linear system has infinitely many solutions. This linear system has number of variables greater than the number of non-zero rows in the REF.*

Note that when solving for linear systems in which the contents are unknown constants, then we need to be careful about performing illegal row operations. That is, assume an augmented matrix

$$\left[\begin{array}{ccc|c} a & 1 & 0 & a \\ 1 & 1 & 1 & 1 \\ 0 & 1 & a & b \end{array} \right] \quad (30)$$

and in order to make the second row leading entry 0, we would perhaps like to perform $R_2 - \frac{1}{a}R_1$. However, we do not know that $a \neq 0$. In this case, we can consider either first swapping the first and second row and progressing, or perform a by-case method.

3.1.1.4 Homogeneous Linear Systems

Definition 18 (Homogeneous Linear Systems). *A linear system (Definition 18) is homogeneous (HLS) if it has augmented matrix representation $(A|b)$ where $b = 0$ and all constants $a_{ij} \in \mathbb{R}, \forall i \in [m], \forall j \in [n]$.*

See that the HLS is always satisfied by $x_i = 0, i \in [n]$ and we call this the trivial (sometimes, zero) solution. A non-trivial solution is any other solution that is not trivial.

Exercise 6. *See that in the xy -plane, the equations*

$$a_1x + b_1y = 0, \quad (31)$$

$$a_2x + b_2y = 0 \quad (32)$$

where a_1, b_1 not both zero and a_2, b_2 not both zero each represent straight lines through the origin, The system has only the trivial solution when the two equations are not the same line, otherwise they have infinitely many solutions. In the xyz -space, a system of two such linear equations passing through the origin always has infinitely many (non-trivial) solutions in addition to the trivial one, since they are either the same plane or intersect at a line passing through the origin at $(0, 0, 0)$.

Lemma 2. *A HLS (Definition 18) has either only the trivial solution or infinitely many solutions in addition to the trivial solution. A HLS with more unknowns than equations has infinitely many solutions.*

Proof. The first assertion is trivial since the zero solution satisfies it. The second assertion follows from considering the REF of the augmented matrix representation of a HLS with more unknowns than equations, then apply Result 2. \square

Exercise 7. *For a HLS $Ax = 0$ (Definition 18) with non-zero solution, show that the system $Ax = b$ has either no solution or infinitely many solutions.*

Proof. By Theorem 2, a HLS system has no solution, one solution or infinite solutions. But suppose there is some solution u s.t. $Au = b$. Let v be non-zero solution for the HLS s.t. $Av = 0$, $v \neq 0$. Then $A(u+v) = Au + Av = b + 0 = b$, so $u+v$ is solution and $u+v \neq u$. But by Lemma 2, the solution space for $Ax = 0$ must have infinitely many vectors if such a v exists. It follows $Ax = b$ has infinitely many solutions if $\exists u$ s.t. $Au = b$. \square

3.1.2 Matrices

We formally defined augmented matrices in Definition 9. In the earlier theorems, we also referred to generalized matrix representations of numbers. We provide formal definition here.

Definition 19 (Matrix). *A matrix is a rectangular array (or array of arrays) of numbers. The numbers are called entries. The size of a matrix is given by the rectangle's sides, and we say a matrix A is $m \times n$ for m rows and n column matrix. We can denote the entry at the i -th row and j -th coordinate by writing $A_{(i,j)} = a_{ij}$. This is often represented*

$$A = \begin{bmatrix} a_{11} & a_{12} \cdots & a_{1n} \\ a_{21} & a_{22} \cdots & a_{2n} \\ \cdots & \cdots \cdots & \cdots \\ a_{m1} & a_{m2} \cdots & a_{mn} \end{bmatrix}, \quad (33)$$

and for brevity we also denote this $A = (a_{ij})_{m \times n}$, and sometimes we drop the size all together and write $A = (a_{ij})$.

Definition 20. *For brevity, given a matrix A (Definition 19) we refer to its size by using the notation $nrows(A)$ and $ncols(A)$ to indicate the number of rows in A and number of columns in A respectively. That is, A is a matrix size $nrows(A) \times ncols(A)$.*

Definition 21 (Column, Row Matrices/Vectors). *A column matrix (vector) is a matrix with only a single column. A row matrix (vector) is a matrix with only one row.*

Definition 22 (Square Matrix). *A square matrix is a matrix (Definition 19) that is square (number of rows is equivalent to the number of rows). We say $A_{n \times n}$ square matrix is of order n .*

Definition 23 (Diagonal Matrix). *A square matrix A of order n (Definition 22) is diagonal matrix if all entries that are not along the diagonal are zero. That is,*

$$a_{ij} = 0 \quad \text{when } i \neq j. \quad (34)$$

Definition 24 (Scalar Matrix). *A diagonal matrix (Definition 23) is scalar matrix if all diagonal entries are the same, that is*

$$a_{ij} = \begin{cases} 0 & i \neq j \\ c & i = j, \end{cases} \quad (35)$$

for some constant $c \in \mathbb{R}$.

Definition 25 (Identity Matrix). *Scalar matrix (Definition 24) is identity matrix if the diagonals are all one, that is $c = 1$. We often denote this as $\mathbb{1}$. If the size needs to be specified, we add subscript $\mathbb{1}_n$ to indicate order n .*