



Marco Aichmann, BSc

Eine Kleinste-Quadrate-Formulierung des Risk-Budgeting-Ansatzes in der Portfoliooptimierung

MASTERARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

Masterstudium Finanz- und Versicherungsmathematik

eingereicht an der

Technischen Universität Graz

Betreuerin:

Ao.Univ.-Prof. Dipl.-Ing. Dr.techn. Eranda Dragoti-Çela

Institut für Optimierung und Diskrete Mathematik
(Math B)

Graz, März 2015

EIDESSTATTLICHE ERKLÄRUNG

AFFIDAVIT

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Textdokument ist mit der vorliegenden Masterarbeit identisch.

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

Datum/Date

Unterschrift/Signature

Danksagungen

Zunächst möchte ich mich an dieser Stelle bei all denjenigen herzlich bedanken, die mich während meines Studiums und vor allem bei der Anfertigung dieser Masterarbeit unterstützt haben.

Danken möchte ich in erster Linie meiner Betreuerin Frau Prof. Dr. Eranda Dragoti-Çela, die mir das interessante Thema für meine Masterarbeit vorgeschlagen hat und stets ein offenes Ohr für meine Anliegen hatte. Vielen Dank für die Bemühungen und die Geduld.

Mein besonderer Dank gilt meiner Mutter und meiner Großmutter, die mir mein Studium ermöglicht haben und mir jederzeit zur Seite standen.

Ich widme diese Arbeit meinem verstorbenen Großvater, der den Abschluss meines Studiums nicht erleben durfte.

Abstract

Im Jahre 1952 veröffentlichte der US-amerikanische Ökonom Harry M. Markowitz ein Paper, welches als Grundlage der modernen Portfoliotheorie angesehen werden kann: Das von ihm eingeführte „Mean-Variance“-Modell beruht auf der Idee, für eine gewisse Anzahl von Assets, in die investiert werden sollte, neben dem Erwartungswert der Returns auch deren Varianz zu berücksichtigen. Dies sollte in einer Art und Weise geschehen, sodass die (angenommene und in der Praxis auch auftretende) Risikoaversität eines Investors ins Modell miteinbezogen wird. Wie revolutionär seine Überlegungen waren, zeigt die Verleihung des Nobelpreises für Wirtschaftswissenschaften, welchen er zusammen mit Merton H. Miller und William F. Sharpe im Jahre 1990 erhielt.

Das von ihm eingeführte Modell hat jedoch neben vielen Vorteilen (wie der äußerst nützlichen Möglichkeit, den Trade-off zwischen erwartetem Return und Portfoliovariabilität zu untersuchen) auch Nachteile, welche selbst von diversen, in den vergangenen 60 Jahren untersuchten Modifikationen und Verbesserungen nicht vollständig beseitigt werden konnten. Es ist beispielsweise bekannt, dass das „Mean-Variance“-Modell recht sensitiv auf Parameteränderungen reagiert.

Diese Arbeit setzt sich zum Ziel, neben dem Markowitz’schen Modell noch eine andere Herangehensweise an die Thematik der Portfoliooptimierung zu untersuchen: das „Risk-Budgeting“-Modell. Der Ansatz des „Risk-Budgeting“ zielt im Allgemeinen darauf ab, in eine Menge von Assets so zu investieren, dass jedes Asset einen bestimmten, vorgegebenen Beitrag zum Gesamtrisiko des Portfolios leistet. Einen besonderen Stellenwert nimmt als Spezialfall das „Risk-Parity“-Modell ein, in dem der Risikobeitrag jedes Assets exakt gleich hoch ist. Die Bezeichnung „Risk-Parity“ wurde dabei erstmals von Qian (2005) verwendet. Als Risikomaß verwenden wir in Analogie zum Markowitz’schen Ansatz die Varianz.

Nach einer Einführung in die Grundlagen der nichtlinearen Optimierung im 1. Kapitel untersuchen wir in den Kapiteln 2 und 3 theoretische Aspekte der beiden erwähnten Portfoliooptimierungsmodelle. Dabei interessieren wir uns vor allem für Formulierungen des „Risk-Budgeting“-Ansatzes durch nichtlineare Optimierungsprobleme. Um in der Lage zu sein, diese zu lösen, betrachten wir im 4. Kapitel verschiedene Optimierungsverfahren. Gegenstand von Kapitel 5, dem praxisbezogenen Teil der Arbeit, sind Implementierungen der entsprechenden Methoden und numerische Berechnungen anhand realer Finanzdaten, um einerseits die Performance unserer Algorithmen bewerten und andererseits Vergleiche zwischen den beiden betrachteten Portfoliooptimierungsmodellen ziehen zu können. Wir interessieren uns dabei im Besonderen für die Zeit der Finanzkrise, die Mitte 2007 begann.

Inhaltsverzeichnis

Abstract	iv
1. Grundlagen der nichtlinearen Optimierung	1
1.1. Allgemeine nichtlineare Optimierungsprobleme	1
1.2. Konvexe Optimierungsprobleme	3
1.3. Quadratische Optimierungsprobleme	5
1.4. Optimalitätsbedingungen	6
1.4.1. Optimalitätsbedingungen für konvexe Optimierungsprobleme	8
1.5. Dualität	11
2. Das „Mean-Variance“-Modell nach Markowitz	15
2.1. Das grundlegende Markowitz’sche Modell	15
2.2. Die effiziente Front	17
2.3. KKT-Bedingungen und äquivalente Modellformulierungen	20
3. Risikobasierte Portfoliooptimierung	27
3.1. Der „Minimum-Variance“-Ansatz	28
3.1.1. Explizite Lösung des „Minimum-Variance“-Problems	29
3.2. Der „Risk-Budgeting“- und „Risk-Parity“-Ansatz	31
3.2.1. Ein Lösungsversuch des „Risk-Parity“-Problems	33
3.3. „Risk-Budgeting“-Portfolios als Lösungen nichtlinearer Optimierungsprobleme	36
3.3.1. Ein konvexes Optimierungsproblem mit logarithmischer Schranke .	37
3.3.2. Eine Kleinste-Quadrate-Formulierung	46
4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme	52
4.1. Quadratische Optimierungsprobleme	53
4.1.1. Behandlung von Problemen mit Gleichheitsrestriktionen	53
4.1.2. Die Methode der aktiven Menge für konvexe Probleme mit Ungleichheitsrestriktionen	55
4.1.3. Innere-Punkte-Verfahren für konvexe Probleme mit Ungleichheitsrestriktionen	65
4.2. Sequentielle Quadratische Optimierung	78
4.2.1. Behandlung von Problemen mit Gleichheitsrestriktionen	78
4.2.2. Eine lokale SQP-Methode für Probleme mit Ungleichheitsrestriktionen	81
4.2.3. Eine globale SQP-Methode für Probleme mit Ungleichheitsrestriktionen	83

Inhaltsverzeichnis

4.2.4. Eine modifizierte BFGS-Formel	91
5. Portfoliooptimierung in der Praxis	95
5.1. Durchgeführte Implementierungen	95
5.2. Numerische Ergebnisse	98
5.2.1. Berechnung von „Risk-Parity“-Portfolios	101
5.2.2. Eine vergleichende Analyse von „Risk-Parity“ und „Mean-Variance“	106
6. Conclusio und Ausblick	123
A. Appendix	126
A.1. Pareto-Optimalität	126
A.2. Positiv homogene Funktionen	127
A.3. Eigenschaften der konstanten Korrelationsmatrix	127
A.4. Cauchy-Schwarz'sche Ungleichung	129
A.5. Abstiegsverfahren	130
A.6. Newton-Methode	132
A.7. Quasi-Newton-Methode	133
Abbildungsverzeichnis	137
Tabellenverzeichnis	138
Literatur	139

1. Grundlagen der nichtlinearen Optimierung

Im ersten Abschnitt möchten wir einige grundlegende Aussagen zu Optimierungsproblemen betrachten. Dieser Teil der Arbeit stammt aus Luenberger und Ye (2008), Cornuéjols und Tütüncü (2007), Nocedal und Wright (2006), Geiger und Kanzow (2002) sowie Geiger und Kanzow (1999).

1.1. Allgemeine nichtlineare Optimierungsprobleme

Sowohl das Markowitz'sche „Mean-Variance“-Modell als auch das „Risk-Budgeting“-Modell, welche wir später näher in verschiedenen Formulierungen untersuchen werden, lassen sich stets als nichtlineare Optimierungsprobleme darstellen. Daher möchten wir zu Beginn die grundlegende Theorie nichtlinearer Programmierung behandeln.

Wir betrachten anfangs folgende generische Form eines Optimierungsproblems:

$$\begin{aligned} \text{(OP)} \quad & \min_{x \in \mathbb{R}^n} f(x) \\ & \text{s.t. } x \in \Omega \end{aligned}$$

mit der Zielfunktion $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ und der Menge $\Omega \subseteq \mathbb{R}^n$ der zulässigen Punkte. Ziel ist es, einen Punkt $x^* \in \mathbb{R}^n$ zu finden, welcher das Problem (OP) löst.

Falls die Menge $\Omega \subseteq \mathbb{R}^n$ leer ist, so heißt das Optimierungsproblem **unzulässig**. Falls eine Sequenz $x^k \in \Omega$ ($k = 1, 2, \dots$) existiert, sodass $f(x^k) \rightarrow -\infty$ (mit $k \rightarrow +\infty$), so nennt man das Problem **unbeschränkt**.

Definition. (*Globales und lokales Minimum*)

Ist (OP) weder unzulässig noch unbeschränkt, so nennt man eine zulässige Lösung x^* von (OP) (d.h. $x^* \in \Omega$)

- **globales Minimum** von (OP), falls $f(x^*) \leq f(x), \forall x \in \Omega$,
- **strikt globales Minimum** von (OP), falls $f(x^*) < f(x), \forall x \in \Omega \setminus \{x^*\}$,
- **lokales Minimum** von (OP), falls $f(x^*) \leq f(x), \forall x \in \Omega \cap B_{x^*}(\varepsilon)$,
- **strikt lokales Minimum** von (OP), falls $f(x^*) < f(x), \forall x \in \{\Omega \setminus \{x^*\}\} \cap B_{x^*}(\varepsilon)$,

1. Grundlagen der nichtlinearen Optimierung

wobei $B_{x^*}(\varepsilon) := \{x \mid \|x - x^*\|_2 < \varepsilon\}$ die offene Kugel in \mathbb{R}^n mit Radius $\varepsilon > 0$ und Mittelpunkt x^* bezeichnet.

Da in vielen Fällen - so auch in den von uns betrachteten Problemen - die Menge der zulässigen Punkte durch Gleichheits- und Ungleichheitsrestriktionen beschrieben wird, beschränken wir uns im Weiteren auf nichtlineare Optimierungsprobleme folgender Form:

$$(NLP) \quad \min_{x \in \mathbb{R}^n} f(x) \quad (1.1)$$

$$s.t. \quad g_i(x) \geq 0, \quad i = 1, \dots, p \quad (1.2)$$

$$h_k(x) = 0, \quad k = 1, \dots, q \quad (1.3)$$

wobei sowohl die Zielfunktion f in (1.1) als auch die Funktionen g_i , $i = 1, \dots, p$ und h_k , $k = 1, \dots, q$ in (1.2) bzw. (1.3) von \mathbb{R}^n nach \mathbb{R} abbilden. Die Restriktionen in (1.2) nennen wir **Ungleichheitsrestriktionen** (inequality constraints), jene in (1.3) **Gleichheitsrestriktionen** (equality constraints).

Damit hat die Menge Ω_{NLP} der zulässigen Punkte für (NLP) die Form

$$\Omega_{NLP} = \{x \in \mathbb{R}^n \mid g_i(x) \geq 0, \forall i = 1, \dots, p \text{ und } h_k(x) = 0, \forall k = 1, \dots, q\}.$$

Weiters werden wir die folgenden Bezeichnungen benötigen:

Definition. (*Aktive Restriktionen*)

Es sei ein nichtlineares Optimierungsproblem (NLP) gegeben.

- Eine Ungleichheitsrestriktion $g_i(x) \geq 0$ heißt **aktiv** in einem zulässigen Punkt $x \in \Omega_{NLP}$, falls $g_i(x) = 0$, und **inaktiv**, falls $g_i(x) > 0$.
- Für einen zulässigen Punkt $x \in \Omega_{NLP}$ ist

$$J(x) := \{i \in \{1, \dots, p\} \mid g_i(x) = 0\}$$

die Menge all jener Indizes, deren zugehörige Ungleichheitsrestriktionen im Punkt x aktiv sind.¹

- Als die „**Menge der** (in einem zulässigen Punkt x) **aktiven Restriktionen**“ bezeichnen wir jene Menge, welche sowohl alle Ungleichheitsrestriktionen $g_i(x)$, die mit Gleichheit erfüllt (also aktiv) sind, als auch alle Gleichheitsrestriktionen $h_k(x)$ beinhaltet.

Für viele Anwendungen sind sogenannte Regularitätsannahmen notwendig. Wir definieren für einen zulässigen Punkt die Regularität der linearen Unabhängigkeit (**linear independence constraint qualification**, LICQ, vgl. Geiger und Kanzow (2002)) nun folgendermaßen:

¹Mit dieser Definition ist also $g_j(x) > 0, \forall j \in \{1, \dots, p\} \setminus J(x)$.

1. Grundlagen der nichtlinearen Optimierung

Definition. (*LICQ-reguläre Punkte*)

Sei $x \in \Omega_{NLP}$ ein *zulässiger* Punkt für (NLP). Dann bezeichnen wir x als **LICQ-regulär** bzgl. des Optimierungsproblems (NLP), falls die Gradientenvektoren $\nabla g_j(x)$ und $\nabla h_k(x)$ mit $j \in J(x)$ und $k \in \{1, \dots, q\}$ linear unabhängig sind.

1.2. Konvexe Optimierungsprobleme

Im folgenden Abschnitt, welcher vollständig auf Geiger und Kanzow (1999) sowie Geiger und Kanzow (2002) basiert, betrachten wir einen Spezialfall der nichtlinearen Optimierung, welcher zu einer signifikanten Vereinfachung der Problemstellung führt. Wir beginnen mit einigen Aussagen zu konvexen Funktionen und Mengen:

Definition. (Konvexe Mengen und Funktionen)

- Eine Menge $M \subseteq \mathbb{R}^n$ heißt **konvex**, falls $\lambda x + (1 - \lambda)y \in M$, $\forall x, y \in M$ und $\forall \lambda \in (0, 1)$.
- Sei $M \subseteq \mathbb{R}^n$ eine nichtleere konvexe Menge. Eine Funktion $f : M \rightarrow \mathbb{R}$ heißt
 - **konvex** auf M , falls

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y),$$

$\forall x, y \in M$ und $\forall \lambda \in (0, 1)$.

- **strikt konvex** auf M , falls

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y),$$

$\forall x, y \in M$ mit $x \neq y$ und $\forall \lambda \in (0, 1)$.

- **(strikt) konkav** auf M , falls die Funktion $-f$ (strikt) konvex auf M ist.

Satz 1.2.1. *Die Menge $M \subseteq \mathbb{R}^n$ sei offen und konvex. Dann ist die Funktion $f : M \rightarrow \mathbb{R}$, für welche wir stetige Differenzierbarkeit annehmen, genau dann konvex auf M , wenn*

$$f(x) - f(y) \geq \nabla f(y)^\top (x - y),$$

$\forall x, y \in M$ gilt, und genau dann strikt konvex auf M , wenn

$$f(x) - f(y) > \nabla f(y)^\top (x - y),$$

$\forall x, y \in M$ mit $x \neq y$ gilt.

Im Rahmen der Konvexität von Funktionen spielt auch die Definitheit von Matrizen eine Rolle:

Definition. (Positiv (semi-)definite Matrizen)

Eine symmetrische Matrix $A \in \mathbb{R}^{n \times n}$ heißt

1. Grundlagen der nichtlinearen Optimierung

- **positiv semidefinit**, falls $x^\top Ax \geq 0, \forall x \in \mathbb{R}^n$. (Notation: $A \succcurlyeq 0$)
- **positiv definit**, falls $x^\top Ax > 0, \forall x \in \mathbb{R}^n \setminus \{0\}$. (Notation: $A \succ 0$)

Satz 1.2.2. *Die Menge $M \subseteq \mathbb{R}^n$ sei offen und konvex. Ist die Funktion $f : M \rightarrow \mathbb{R}$ zweimal stetig differenzierbar, dann ist f genau dann konvex auf M , wenn $\nabla^2 f(x)$ für alle $x \in M$ positiv semidefinit ist. Falls $\nabla^2 f(x)$ für alle $x \in M$ sogar positiv definit ist, so ist f strikt konvex auf M .*

Für Beweise zu den Sätzen 1.2.1 und 1.2.2 siehe Geiger und Kanzow (1999).

Ein konvexes Optimierungsproblem (auch **konvexes Programm** (KP) genannt) hat nun folgende Form:

$$(KP) \quad \min_{x \in \mathbb{R}^n} f(x) \tag{1.4}$$

$$s.t. \quad g_i(x) \geq 0, \quad i = 1, \dots, p \tag{1.5}$$

$$Bx = \beta \tag{1.6}$$

wobei $B \in \mathbb{R}^{q \times n}, \beta \in \mathbb{R}^q$, die Zielfunktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ konvex und die Funktionen $g_i : \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, \dots, p$ konkav sind. Diese Art von Optimierungsproblem ist also ein nichtlineares Optimierungsproblem mit konvexer Zielfunktion und konkaven Ungleichheitsrestriktionen sowie linearen Gleichheitsrestriktionen.

Schreiben wir die Restriktionen (1.5) als $g'_i(x) \leq 0$ mit $g'_i = -g_i, \forall i = 1, \dots, p$, so ist zu fordern, dass die Funktionen $g'_i : \mathbb{R}^n \rightarrow \mathbb{R}$ (so wie auch die Zielfunktion f) konvex sind. Diese Variante der Modelldefinition ist oftmals in der Literatur zu finden und rechtfertigt die Bezeichnung von (KP) als *konvexes* Optimierungsproblem.

Die Menge der zulässigen Punkte $\Omega_{KP} = \{x \in \mathbb{R}^n \mid g_i(x) \geq 0, \forall i = 1, \dots, p \text{ und } Bx = \beta\}$ ist hier eine konvexe Menge (siehe Geiger und Kanzow, 2002). Daher ist folgender Satz für konvexe Optimierungsprobleme von Bedeutung:

Satz 1.2.3. *Betrachte das generische Optimierungsproblem (OP). Die Menge der zulässigen Punkte $\Omega \subseteq \mathbb{R}^n$ sei nichtleer und konvex. Weiters sei die Zielfunktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ konvex (bzw. strikt konvex). Dann ist jedes lokale Minimum von (OP) auch ein globales (bzw. strikt globales) Minimum.*

Beweis. Folgt unmittelbar aus der Konvexität von Ω und der (strikten) Konvexität von f (vgl. Geiger und Kanzow, 2002). □

1.3. Quadratische Optimierungsprobleme

Im Rahmen dieser Arbeit werden quadratische Optimierungsprobleme - auch **quadratische Programme** (QP) genannt - eine wichtige Rolle spielen. So werden wir später beispielsweise Formulierungen des Markowitz'schen „Mean-Variance“-Modells als quadratische Programme kennenlernen. Daher betrachten wir an dieser Stelle noch diesen Spezialfall eines nichtlinearen Optimierungsproblems, in welchem die Zielfunktion eine quadratische Funktion der Variablen ist und die Restriktionen in linearer Form vorliegen:

$$\begin{aligned}
 \text{(QP)} \quad & \min_{x \in \mathbb{R}^n} \quad \frac{1}{2} x^\top Q x + c^\top x \\
 & \text{s.t.} \quad Ax \geq \alpha \\
 & \quad \quad Bx = \beta
 \end{aligned}$$

wobei $Q \in \mathbb{R}^{n \times n}$, $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{p \times n}$, $\alpha \in \mathbb{R}^p$, $B \in \mathbb{R}^{q \times n}$ und $\beta \in \mathbb{R}^q$ gegeben sind. Da $x^\top Q x \in \mathbb{R}^1$ und somit $x^\top Q x = (x^\top Q x)^\top = x^\top Q^\top x$ gilt, folgt $x^\top Q x = \frac{1}{2} x^\top (Q + Q^\top) x$, weshalb wir ohne Beschränkung der Allgemeinheit annehmen können, dass die Matrix Q symmetrisch ist.

Mithilfe von Satz 1.2.2 und der Tatsache, dass die Hesse-Matrix $\nabla^2 (\frac{1}{2} x^\top Q x + c^\top x) = Q$ ist (und zwar für alle zulässigen Punkte x) erhalten wir nun folgende Aussage:

Falls Q eine positiv semidefinite Matrix ist ($Q \succcurlyeq 0$), dann ist die Zielfunktion des Problems (QP) konvex. Ist Q sogar positiv definit ($Q \succ 0$), so erhalten wir eine strikt konvexe Zielfunktion. Offensichtlich haben wir also in diesem Fall (positiv semidefinite Matrix Q) wiederum einen Spezialfall des konvexen Programms (KP), da sowohl die Gleichheits- als auch die Ungleichheitsrestriktionen in (QP) in linearer (und somit auch konkaver) Form vorliegen.

Des Weiteren werden wir auch die folgende zu (QP) äquivalente Darstellung eines quadratischen Optimierungsproblems mit $\tilde{Q} \in \mathbb{R}^{\tilde{n} \times \tilde{n}}$, $\tilde{c} \in \mathbb{R}^{\tilde{n}}$, $\tilde{B} \in \mathbb{R}^{\tilde{q} \times \tilde{n}}$ und $\tilde{\beta} \in \mathbb{R}^{\tilde{q}}$ verwenden:

$$\begin{aligned}
 \text{(\tilde{Q}P)} \quad & \min_{\tilde{x} \in \mathbb{R}^{\tilde{n}}} \quad \frac{1}{2} \tilde{x}^\top \tilde{Q} \tilde{x} + \tilde{c}^\top \tilde{x} \\
 & \text{s.t.} \quad \tilde{B} \tilde{x} = \tilde{\beta} \\
 & \quad \quad \tilde{x} \geq 0
 \end{aligned}$$

Dass jedes Programm (QP) in die Form $(\tilde{Q}P)$ gebracht werden kann, ist leicht einzusehen:

Man führt in (QP) eine neue Variable $y \in \mathbb{R}^p$ mit $y \geq 0$ ein und ersetzt die Ungleichheitsrestriktionen durch $Ax - y = \alpha$. Darüber hinaus verwendet man $x = x^+ - x^-$ mit $x_i^+ := \max(x_i, 0) \geq 0$ und $x_i^- := -\min(x_i, 0) \geq 0$ und setzt schlussendlich

$$\tilde{x} := \begin{pmatrix} x^+ \\ x^- \end{pmatrix}, \quad \tilde{Q} := \begin{pmatrix} Q & -Q & 0 \\ -Q & Q & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \tilde{c} := \begin{pmatrix} c \\ 0 \end{pmatrix}, \quad \tilde{B} := \begin{pmatrix} A & -A & -I_p \\ B & -B & 0 \end{pmatrix}, \quad \tilde{\beta} := \begin{pmatrix} \alpha \\ \beta \end{pmatrix},$$

sowie $\tilde{n} := 2n + p$ und $\tilde{q} := p + q$. Bei dieser Transformation sind die folgenden beiden Beobachtungen für unsere späteren Untersuchungen von Bedeutung:

1. Grundlagen der nichtlinearen Optimierung

- Ist die Matrix Q positiv definit, so ist \tilde{Q} ebenfalls positiv definit, falls $p = 0$, und positiv semidefinit, falls $p > 0$.
- Hat die Matrix B vollen Zeilenrang ($\text{rank}(B) = q$), so gilt dies auch für \tilde{B} ($\text{rank}(\tilde{B}) = \tilde{q}$).

1.4. Optimalitätsbedingungen

Wir interessieren uns nun für Bedingungen, welche für das Vorliegen eines lokalen Minimums bei einem nichtlinearen Optimierungsproblem notwendig bzw. hinreichend sind. Diese in den folgenden drei Sätzen formulierten Verallgemeinerungen der Lagrange-Multiplikatoren werden oftmals auch **Karush-Kuhn-Tucker** (KKT) Bedingungen genannt.

In diesem Teil der Arbeit nehmen wir an, dass sowohl die Zielfunktion als auch die Gleichheits- und Ungleichheitsrestriktionen so oft stetig differenzierbar sind wie nötig.

Satz 1.4.1. (*KKT Bedingungen - Notwendige Bedingungen 1. Ordnung*)

Sei $x^* \in \Omega_{NLP}$ ein lokales Minimum und LICQ-regulär für (NLP). Dann existieren Skalare λ_i , $i = 1, \dots, p$ und γ_k , $k = 1, \dots, q$, sodass

$$\nabla f(x^*) - \sum_{i=1}^p \lambda_i \nabla g_i(x^*) - \sum_{k=1}^q \gamma_k \nabla h_k(x^*) = 0 \quad (1.7)$$

sowie

$$\lambda_i \geq 0, \quad i = 1, \dots, p \quad (1.8)$$

$$\lambda_i \cdot g_i(x^*) = 0, \quad i = 1, \dots, p \quad (1.9)$$

gelten.

Beweis. Siehe Luenberger und Ye (2008) sowie Nocedal und Wright (2006). □

Fassen wir die obigen Skalare in Vektoren $\lambda = (\lambda_1, \dots, \lambda_p)^\top$ und $\gamma = (\gamma_1, \dots, \gamma_q)^\top$ zusammen und definieren wir die sogenannte **Lagrangefunktion**

$$\mathcal{L}(x, \lambda, \gamma) := f(x) - \sum_{i=1}^p \lambda_i g_i(x) - \sum_{k=1}^q \gamma_k h_k(x), \quad (1.10)$$

so entspricht die linke Seite von Ausdruck (1.7) genau dem Gradienten $\nabla_x \mathcal{L}(x, \lambda, \gamma)|_{x=x^*}$ (mit den partiellen Ableitungen nach x_i , $i = 1, \dots, n$).

Die Voraussetzungen in (1.9) werden (in Kombination mit (1.8)) auch als die Bedingungen vom komplementären Schlupf bezeichnet und implizieren, dass der Lagrange-Multiplikator λ_i gleich 0 sein muss, falls die i -te Ungleichheitsrestriktion *inaktiv* ist.

1. Grundlagen der nichtlinearen Optimierung

Definition. (KKT-Punkt)

Ein Vektor $(x^*, \lambda, \gamma) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q$ heißt **KKT-Punkt** für das Optimierungsproblem (NLP), falls er die Bedingungen (1.7), (1.8) sowie (1.9) erfüllt und x^* zulässig (also $x^* \in \Omega_{NLP}$) ist.

Man beachte, dass die LICQ-Regularität des Punktes x^* *nicht* Teil der Definition eines KKT-Punktes ist.

Oftmals treten in der Praxis nichtlineare Optimierungsprobleme (NLP) auf, deren Gleichheits- und Ungleichheitsrestriktionen in **linearer** Form vorliegen. Für diesen Spezialfall ist folgendes Korollar hilfreich:

Korollar 1.4.2. *Satz 1.4.1 bleibt korrekt, falls die angenommene LICQ-Regularität von x^* durch die Annahme der Linearität der Funktionen g_i und h_k ersetzt wird.*

Beweis. Siehe Geiger und Kanzow (2002). □

Für den Fall, dass sämtliche Restriktionen linear sind, ist also jedes lokale Minimum x^* ein KKT-Punkt, ohne dass weitere Regularitätsannahmen an x^* gestellt werden müssen.

Neben lokalen Minima ist Satz 1.4.1 auch für lokale Maxima sowie Sattelpunkte erfüllt, weshalb wir noch Bedingungen 2. Ordnung betrachten möchten. Die Formulierung der folgenden beiden Sätze entspricht jener in Cornuéjols und Tütüncü (2007):

Satz 1.4.3. *(KKT Bedingungen - Notwendige Bedingungen 2. Ordnung)*

Sei $x^ \in \Omega_{NLP}$ ein lokales Minimum und LICQ-regulär für (NLP).*

$\mathcal{J}_{act}(x^)$ bezeichne die Jacobi-Matrix der Menge der in x^* aktiven Restriktionen und $B_{act}(x^*)$ sei eine Basis des Kerns von $\mathcal{J}_{act}(x^*)$.*

Dann existieren Skalare λ_i , $i = 1, \dots, p$ und γ_k , $k = 1, \dots, q$, sodass die Bedingungen (1.7), (1.8) und (1.9) sowie

$$B_{act}^\top(x^*) \left(\nabla^2 f(x^*) - \sum_{i=1}^p \lambda_i \nabla^2 g_i(x^*) - \sum_{k=1}^q \gamma_k \nabla^2 h_k(x^*) \right) B_{act}(x^*) \succcurlyeq 0$$

erfüllt sind.

Auch dieser Satz liefert nur eine weitere notwendige Bedingung für das Vorliegen eines Minimums. Mithilfe der Informationen 2. Ordnung lässt sich jedoch auch folgender Satz formulieren, welcher hinreichende Kriterien angibt:

Satz 1.4.4. *(KKT Bedingungen - Hinreichende Bedingungen 2. Ordnung)*

Sei $x^ \in \Omega_{NLP}$ LICQ-regulär für (NLP).*

$\mathcal{J}_{act}(x^)$ bezeichne die Jacobi-Matrix der Menge der in x^* aktiven Restriktionen und $B_{act}(x^*)$ sei eine Basis des Kerns von $\mathcal{J}_{act}(x^*)$.*

1. Grundlagen der nichtlinearen Optimierung

Falls Skalare λ_i , $i = 1, \dots, p$ und γ_k , $k = 1, \dots, q$ existieren, sodass die Bedingungen (1.7), (1.8) und (1.9) sowie

$$B_{act}^\top(x^*) \left(\nabla^2 f(x^*) - \sum_{i=1}^p \lambda_i \nabla^2 g_i(x^*) - \sum_{k=1}^q \gamma_k \nabla^2 h_k(x^*) \right) B_{act}(x^*) \succ 0$$

und

$$g_i(x^*) = 0, \quad i \in \{1, \dots, p\} \Rightarrow \lambda_i > 0$$

erfüllt sind, dann ist x^* ein strikt lokales Minimum von (NLP).

Alternative Formulierungen sowie Beweise der Sätze (1.4.3) und (1.4.4) sind in Luenberger und Ye (2008) sowie Nocedal und Wright (2006) zu finden.

1.4.1. Optimalitätsbedingungen für konvexe Optimierungsprobleme

Unter gewissen Bedingungen an die Zielfunktion sowie die Restriktionen sind jedoch bereits die KKT-Bedingungen 1. Ordnung hinreichend für das Vorliegen eines *globalen* Minimums. Dies ist beispielsweise für ein konvexes Optimierungsproblem der Fall.

Wir formulieren nochmals die KKT-Bedingungen 1. Ordnung für den Spezialfall konvexer Programme. Dafür verwenden wir folgende Schreibweise für die Matrix B und den Vektor β , welche in (KP) die Gleichheitsrestriktionen definieren:

$$B = \begin{pmatrix} b_1^\top \\ b_2^\top \\ \vdots \\ b_q^\top \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{pmatrix}$$

mit $b_k \in \mathbb{R}^n$, $\forall k = 1, \dots, q$.

Mit dieser Notation lässt sich das Problem (KP) schreiben als:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \geq 0, \quad i = 1, \dots, p \\ & b_k^\top x = \beta_k, \quad k = 1, \dots, q \end{aligned}$$

Satz 1.4.5. (KKT Bedingungen 1. Ordnung für (KP))

Sei $x^* \in \Omega_{KP}$ ein lokales Minimum und LICQ-regulär für (KP). Dann existieren Skalare λ_i , $i = 1, \dots, p$ und γ_k , $k = 1, \dots, q$, sodass

$$\nabla f(x^*) - \sum_{i=1}^p \lambda_i \nabla g_i(x^*) - \sum_{k=1}^q \gamma_k b_k = \nabla f(x^*) - \sum_{i=1}^p \lambda_i \nabla g_i(x^*) - B^\top \gamma = 0 \quad (1.11)$$

1. Grundlagen der nichtlinearen Optimierung

sowie

$$\lambda_i \geq 0, \quad i = 1, \dots, p \quad (1.12)$$

$$\lambda_i \cdot g_i(x^*) = 0, \quad i = 1, \dots, p \quad (1.13)$$

gelten.

Beweis. Folgt direkt aus Satz (1.4.1) mit $h_k(x^*) = b_k^\top x^* - \beta_k$ und somit $\nabla h_k(x^*) = b_k$. \square

Obiger Satz bleibt richtig, wenn anstatt der LICQ-Regularität von x^* folgende Bedingung an das Problem (KP) gestellt wird:

Definition. (Regularitätsbedingung von Slater (vgl. Jarre und Stoer, 2004))

Ein konvexes Optimierungsproblem (KP) erfüllt die **Regularitätsbedingung von Slater** oder ist **Slater-regulär**, wenn ein Punkt $\hat{x} \in \Omega_{KP}$ (also ein *zulässiger* Punkt mit $g_i(\hat{x}) \geq 0$, $\forall i = 1, \dots, p$ und $B\hat{x} = \beta$) existiert, sodass $\forall i \in \{1, \dots, p\}$ gilt:

$$g_i \text{ ist nichtlinear} \Rightarrow g_i(\hat{x}) > 0.$$

Es muss also einen zulässigen Punkt geben, welcher alle *nichtlinearen* Ungleichheitsrestriktionen strikt erfüllt.

Diese Bedingung ist auch ohne Kenntnis der Optimallösung verifizierbar. Wir erhalten also für konvexe Optimierungsprobleme folgende Aussage mit oftmals einfacher überprüfbareren Regularitätsannahmen:

Satz 1.4.6. (KKT Bedingungen 1. Ordnung für (KP) unter Slater-Regularität)

Erfülle das konvexe Optimierungsproblem (KP) die Regularitätsbedingung von Slater und sei $x^* \in \Omega_{KP}$ ein lokales Minimum für (KP). Dann existieren Skalare λ_i , $i = 1, \dots, p$ und γ_k , $k = 1, \dots, q$, sodass (1.11), (1.12) und (1.13) gelten.

Beweis. Siehe Geiger und Kanzow (2002) für den (etwas restriktiveren und leichter zu beweisenden) Fall, dass ein zulässiger Punkt existiert, welcher *alle* Ungleichheitsrestriktionen - also auch die linearen - strikt erfüllt. \square

Wir werden später eine Darstellung des „Mean-Variance“-Modells kennenlernen, für welche wir diese Art der Regularität benötigen werden.

Es lässt sich nun folgender Satz formulieren:

Satz 1.4.7. Sei $(x^*, \lambda, \gamma) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q$ ein KKT-Punkt des konvexen Optimierungsproblems (KP). Dann ist x^* ein lokales (und somit auch ein globales) Minimum von (KP).

1. Grundlagen der nichtlinearen Optimierung

Beweis. (Aus Geiger und Kanzow, 2002.)

Wir betrachten einen beliebigen zulässigen Punkt $x \in \Omega_{KP}$. Da (x^*, λ, γ) ein KKT-Punkt und die Zielfunktion f laut Annahme konvex ist, erhalten wir

$$\begin{aligned} f(x) &\stackrel{\substack{\text{Satz} \\ 1.2.1}}{\geq} f(x^*) + \nabla f(x^*)^\top (x - x^*) \\ &\stackrel{(1.11)}{=} f(x^*) + \sum_{i=1}^p \lambda_i \nabla g_i(x^*)^\top (x - x^*) + \sum_{k=1}^q \gamma_k b_k^\top (x - x^*) \\ &= f(x^*) + \sum_{j \in J(x^*)} \lambda_j \nabla g_j(x^*)^\top (x - x^*), \end{aligned}$$

wobei $J(x^*)$ jene Indizes beinhaltet, deren zugehörige Ungleichheitsrestriktionen in x^* aktiv sind.

Die letzte Gleichung folgt aus der Tatsache, dass einerseits aufgrund der komplementären Schlupfbedingungen in (1.12) und (1.13) die Lagrange-Multiplikatoren λ_i für alle in x^* *inaktiven* Ungleichheitsrestriktionen g_i gleich 0 sein müssen und andererseits wegen der Zulässigkeit von x und x^*

$$\sum_{k=1}^q \gamma_k b_k^\top (x - x^*) = \sum_{k=1}^q \gamma_k (b_k^\top x - b_k^\top x^*) = \sum_{k=1}^q \gamma_k (\beta_k - \beta_k) = 0$$

gilt.

Aufgrund der Konkavität der g_j und $\lambda_j \geq 0$ folgt nun für die in x^* aktiven Ungleichheitsrestriktionen (durch Anwendung von Satz 1.2.1 auf die konvexen Funktionen $-g_j$), dass

$$\lambda_j \nabla g_j(x^*)^\top (x - x^*) \geq \lambda_j (g_j(x) - g_j(x^*)) = \lambda_j g_j(x) \geq 0$$

für alle $j \in J(x^*)$. Daraus folgt insgesamt $f(x) \geq f(x^*)$ und damit die lokale (und mit Satz 1.2.3 auch globale) Optimalität von x^* für (KP). \square

Es gilt zu beachten, dass es für den obigen Satz *nicht* notwendig ist, Regularitätsbedingungen an den Punkt x^* zu stellen.

Mithilfe von Korollar 1.4.2 und den obigen Ergebnissen für konvexe Programme lässt sich nun folgendes Korollar formulieren, welches beispielsweise auf quadratische Programme (QP) mit positiv semidefiniter Matrix Q anwendbar und daher für uns relevant ist:

Korollar 1.4.8. *Betrachte ein konvexes Optimierungsproblem der Form (KP), welches neben linearen Gleichheits- auch **lineare** Ungleichheitsrestriktionen besitzt. Dann ist $x^* \in \mathbb{R}^n$ ein lokales (und damit auch globales) Minimum von (KP), genau dann wenn es Vektoren $\lambda \in \mathbb{R}^p$ und $\gamma \in \mathbb{R}^q$ gibt, sodass (x^*, λ, γ) ein KKT-Punkt von (KP) ist.*

1.5. Dualität

Wir werden im 4. Kapitel Algorithmen zur Lösung nichtlinearer Optimierungsprobleme untersuchen und dabei unter anderem für konvexe quadratische Programme sogenannte Innere-Punkte-Verfahren verwenden. Da in diesem Zusammenhang auch die Dualitätstheorie eine Rolle spielt, möchten wir dieses Thema nun kurz anschnitten. Sämtliche Resultate stammen aus Nocedal und Wright (2006).

Wir betrachten ein nichtlineares Optimierungsproblem (NLP) mit der zugehörigen, in (1.10) definierten Lagrangefunktion, wobei wir (wie im letzten Abschnitt) wieder voraussetzen, dass die Zielfunktion und die Restriktionen so oft stetig differenzierbar sind wie nötig. Des Weiteren definieren wir die Lagrange-duale Zielfunktion \mathcal{L}^* :

$$\begin{aligned} \mathcal{L}^* : \mathbb{R}^p \times \mathbb{R}^q &\rightarrow \mathbb{R} \\ (\lambda, \gamma) &\mapsto \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \gamma) \end{aligned}$$

Dann ist das **duale** oder **Lagrange-duale** Problem von (NLP) folgendermaßen definiert:

$$\begin{aligned} \text{(NLP)}^{\text{L-d}} \quad & \max_{\substack{\lambda \in \mathbb{R}^p \\ \gamma \in \mathbb{R}^q}} \mathcal{L}^*(\lambda, \gamma) \\ & \text{s.t. } \lambda \geq 0 \end{aligned}$$

Man kann sich dabei auf die Menge $\mathcal{D} := \{(\lambda, \gamma) \in \mathbb{R}^p \times \mathbb{R}^q \mid \mathcal{L}^*(\lambda, \gamma) > -\infty\}$ als Definitionsmenge von \mathcal{L}^* beschränken, welche nur jene Werte (λ, γ) beinhaltet, für die \mathcal{L}^* endlich ist.

Zwischen dem Problem (NLP), welches als *primales* Problem bezeichnet wird, und dem dualen Problem $(\text{NLP})^{\text{L-d}}$ bestehen interessante Zusammenhänge, wobei diese für konvexe Programme (KP) noch stärker sind. Zur einfacheren Formulierung werden wir folgende Bezeichnung verwenden:

Definition. (Reguläre konvexe Optimierungsprobleme)

Es seien ein konvexes Optimierungsproblem (KP) und ein zulässiger Punkt $x \in \mathbb{R}^n$ von (KP) gegeben. Dann sagen wir, dass „(KP) **bzw.** x **einer hinreichenden Regularitätsbedingung genügt**“, falls zumindest eine der drei folgenden Voraussetzungen erfüllt ist:

- Alle Ungleichheitsrestriktionen von (KP) sind linear.
- Das Problem (KP) erfüllt die Regularitätsbedingung von Slater.
- Der Punkt x ist LICQ-regulär für (KP).

Es gelten folgende Eigenschaften:

1. Grundlagen der nichtlinearen Optimierung

Satz 1.5.1. (Lagrange-Dualität)

- a) Die Funktion $\mathcal{L}^*(\lambda, \gamma)$ ist konkav in (λ, γ) .
- b) Die Definitionsmenge \mathcal{D} ist konvex.
- c) **Schwache Dualität:**
Für jeden zulässigen Punkt x von (NLP) und jeden zulässigen Punkt (λ, γ) von (NLP)^{L-d} gilt $\mathcal{L}^*(\lambda, \gamma) \leq f(x)$.
- d) **Starke Dualität:**
Es seien ein konvexes Optimierungsproblem (KP) und eine Optimallösung $x^* \in \mathbb{R}^n$ von (KP) gegeben.
 - i) Jeder Punkt (λ^*, γ^*) , für den $(x^*, \lambda^*, \gamma^*)$ ein KKT-Punkt von (KP) ist, ist eine Lösung des Lagrange-dualen Problems (KP)^{L-d}.
 - ii) (KP) bzw. x^* genüge einer hinreichenden Regularitätsbedingung und $(\hat{\lambda}, \hat{\gamma})$ sei eine Lösung des Lagrange-dualen Problems (KP)^{L-d}. Ist $\mathcal{L}(x, \hat{\lambda}, \hat{\gamma})$ strikt konvex in x und $\hat{x} = \operatorname{arginf}_{x \in \mathbb{R}^n} \mathcal{L}(x, \hat{\lambda}, \hat{\gamma})$, so gilt $x^* = \hat{x}$ und $f(x^*) = \mathcal{L}(\hat{x}, \hat{\lambda}, \hat{\gamma})$.

Die Beweise dieser Aussagen sind in Nocedal und Wright (2006) nachzulesen. Man beachte in d)-i), dass die Existenz eines KKT-Punktes garantiert ist, falls (KP) bzw. x^* einer hinreichenden Regularitätsbedingung genügt. Des Weiteren ist in d)-ii) die Funktion $\mathcal{L}(x, \hat{\lambda}, \hat{\gamma})$ als Summe von konvexen Funktionen stets konvex in x . Wenn die Zielfunktion f oder g_i für zumindest ein $i \in \{1, \dots, p\}$ mit $\hat{\lambda}_i > 0$ strikt konvex ist, dann ist auch $\mathcal{L}(x, \hat{\lambda}, \hat{\gamma})$ strikt konvex in x .

Wir betrachten nun noch das Konzept der **Wolfe-Dualität** etwas genauer. Das **Wolfe-duale** Problem von (NLP) hat folgende Form:

$$\begin{aligned}
 \text{(NLP)}^{\text{W-d}} \quad & \max_{\substack{x \in \mathbb{R}^n \\ \lambda \in \mathbb{R}^p \\ \gamma \in \mathbb{R}^q}} \mathcal{L}(x, \lambda, \gamma) \\
 & \text{s.t.} \quad \nabla_x \mathcal{L}(x, \lambda, \gamma) = 0 \\
 & \quad \lambda \geq 0
 \end{aligned}$$

Es besteht nun für konvexe Optimierungsprobleme nachstehender Zusammenhang zwischen dem primalen Problem (KP) und dem Wolfe-dualen Problem (KP)^{W-d}:

Satz 1.5.2. (Wolfe-Dualität)

Es seien ein konvexes Optimierungsproblem (KP) und eine Optimallösung $x^* \in \mathbb{R}^n$ von (KP) gegeben. Dann ist jeder KKT-Punkt $(x^*, \lambda^*, \gamma^*)$ von (KP) auch eine Lösung des Wolfe-dualen Problems (KP)^{W-d}.

Beweis. (Vgl. Nocedal und Wright, 2006.)

Da $(x^*, \lambda^*, \gamma^*)$ laut Annahme ein KKT-Punkt ist, gelten die Bedingungen $\lambda^* \geq 0$ sowie

1. Grundlagen der nichtlinearen Optimierung

$\mathcal{L}(x^*, \lambda^*, \gamma^*) = f(x^*)$ und $\nabla_x \mathcal{L}(x, \lambda^*, \gamma^*)|_{x=x^*} = 0$, womit in diesem Punkt die Restriktionen von (KP)^{W-d} erfüllt sind. Daher gilt für jeden beliebigen (für (KP)^{W-d} zulässigen) Punkt (x, λ, γ) , dass

$$\begin{aligned} \mathcal{L}(x^*, \lambda^*, \gamma^*) &= f(x^*) \\ &\geq f(x^*) - \sum_{i=1}^p \lambda_i g_i(x^*) - \gamma^\top (Bx^* - \beta) \\ &= \mathcal{L}(x^*, \lambda, \gamma) \\ &\geq \mathcal{L}(x, \lambda, \gamma) + \nabla_x \mathcal{L}(x, \lambda, \gamma)^\top (x^* - x) \\ &= \mathcal{L}(x, \lambda, \gamma) \end{aligned}$$

Die zweite Ungleichung folgt aus der Tatsache, dass die Funktion $\mathcal{L}(x, \lambda, \gamma)$ (wie bereits zuvor erwähnt) stets konvex in x ist. Die letzte Gleichung ist ein Resultat der dualen Zulässigkeit von (x, λ, γ) , da diese $\nabla_x \mathcal{L}(x, \lambda, \gamma) = 0$ impliziert.

Damit ist $(x^*, \lambda^*, \gamma^*)$ ein Maximum von (KP)^{W-d}. □

Die Existenz eines KKT-Punktes in obigem Satz ist wieder sichergestellt, falls (KP) bzw. x^* einer hinreichenden Regularitätsbedingung genügt.

Dualität und konvexe quadratische Optimierungsprogramme

Wir ermitteln nun die dualen Probleme des quadratischen Optimierungsproblems (QP) aus Abschnitt 1.3. Als Lagrange-duale Zielfunktion erhalten wir

$$\mathcal{L}^*(\lambda, \gamma) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \gamma) = \inf_{x \in \mathbb{R}^n} \frac{1}{2} x^\top Q x + c^\top x - \lambda^\top (Ax - \alpha) - \gamma^\top (Bx - \beta).$$

Falls die Matrix Q positiv definit ist, so ist $\mathcal{L}(x, \lambda, \gamma)$ strikt konvex in x und das eindeutige Infimum wird in $\nabla_x \mathcal{L}(x, \lambda, \gamma) = Qx + c - A^\top \lambda - B^\top \gamma = 0$ angenommen. Da Q wegen $Q \succ 0$ invertierbar ist (und o.B.d.A. symmetrisch), so folgt $x = Q^{-1} (A^\top \lambda + B^\top \gamma - c)$ und somit insgesamt für das Lagrange-duale Programm folgende Darstellung:

$$\begin{aligned} \text{(QP)}^{\text{L-d}} \quad & \max_{\substack{\lambda \in \mathbb{R}^p \\ \gamma \in \mathbb{R}^q}} & -\frac{1}{2} (A^\top \lambda + B^\top \gamma - c)^\top Q^{-1} (A^\top \lambda + B^\top \gamma - c) + \alpha^\top \lambda + \beta^\top \gamma \\ & \text{s.t.} \quad & \lambda \geq 0 \end{aligned}$$

Im Wolfe-dualen Programm werden dagegen die Variablen x_1, \dots, x_n nicht eliminiert. Stattdessen verwenden wir

$$Qx + c - A^\top \lambda - B^\top \gamma = 0 \quad \Rightarrow \quad (c - A^\top \lambda - B^\top \gamma)^\top x = -x^\top Qx$$

1. Grundlagen der nichtlinearen Optimierung

womit das Wolfe-duale Problem von (QP) wie folgt angegeben werden kann:

$$\begin{aligned} (\text{QP})^{\text{W-d}} \quad & \max_{\substack{x \in \mathbb{R}^n \\ \lambda \in \mathbb{R}^p \\ \gamma \in \mathbb{R}^q}} -\frac{1}{2}x^\top Qx + \alpha^\top \lambda + \beta^\top \gamma \\ & \text{s.t.} \quad Qx + c - A^\top \lambda - B^\top \gamma = 0 \\ & \quad \lambda \geq 0 \end{aligned}$$

Für die Herleitung von $(\text{QP})^{\text{W-d}}$ wurde die positive Definitheit von Q nicht verwendet. Um Satz 1.5.2 anwenden zu können, ist es somit ausreichend, wenn $Q \succcurlyeq 0$ ist.

Wir geben nun noch das Wolfe-duale Problem des ebenfalls in Abschnitt 1.3 definierten Programms $(\widetilde{\text{QP}})$ an, welches wir für die Innere-Punkte-Verfahren benötigen werden. Dieses lässt sich unmittelbar aus $(\text{QP})^{\text{W-d}}$ ablesen:

$$\begin{aligned} (\widetilde{\text{QP}})^{\text{W-d}} \quad & \max_{\substack{\tilde{x} \in \mathbb{R}^{\tilde{n}} \\ \tilde{\gamma} \in \mathbb{R}^{\tilde{q}}}} -\frac{1}{2}\tilde{x}^\top \tilde{Q}\tilde{x} + \tilde{\beta}^\top \tilde{\gamma} \\ & \text{s.t.} \quad \tilde{Q}\tilde{x} + \tilde{c} - \tilde{\lambda} - \tilde{B}^\top \tilde{\gamma} = 0 \\ & \quad \tilde{\lambda} \geq 0 \end{aligned}$$

2. Das „Mean-Variance“-Modell nach Markowitz

Dieses Kapitel basiert auf Markowitz (1952), Markowitz (1987), Steinbach (2001), Cornuéjols und Tütüncü (2007) sowie Brito und Vicente (2012).

2.1. Das grundlegende Markowitz'sche Modell

Wir beginnen mit der Definition des grundlegenden Markowitz'schen Modells, wobei wir uns weitgehend an Cornuéjols und Tütüncü (2007) sowie Brito und Vicente (2012) halten.

Es sei eine Menge von n Assets gegeben. Der Return des i -ten Assets (in einem bestimmten Zeitraum $[T_0, T_1]$) wird durch die Zufallsvariable $R_i = \frac{P_i^{T_1} - P_i^{T_0}}{P_i^{T_0}}$ beschrieben, wobei $P_i^{T_0}$ dem (bekannten) Preis des i -ten Assets zum Zeitpunkt T_0 und $P_i^{T_1}$ dem (unbekannten, zufälligen) Preis des i -ten Assets zum Zeitpunkt T_1 entspricht. Wir bezeichnen mit $\mu_i := \mathbb{E}(R_i)$, $i = 1, \dots, n$ die erwarteten Returns der n Assets.

Ein Portfolio wird nun durch Gewichte w_i , $i = 1, \dots, n$ beschrieben, wobei w_i jener *Anteil* der gesamten Investitionssumme ist, welcher in das i -te Asset investiert wird. Dabei wird angenommen, dass der Return des Portfolios linear in w_1, \dots, w_n ist. Fassen wir die Gewichte in einem Gewichtsvektor $w = (w_1, \dots, w_n)^\top$ zusammen, so beschreibt die Zufallsvariable $R(w) := w_1 R_1 + \dots + w_n R_n$ also den Return eines durch w_1, \dots, w_n definierten Portfolios. Aufgrund der Definition der Gewichte als Anteile an der gesamten Investitionssumme muss die Bedingung

$$\sum_{i=1}^n w_i = e^\top w = 1$$

mit $e = (1, \dots, 1)^\top \in \mathbb{R}^n$ erfüllt sein.

Für den erwarteten Return des gesamten Portfolios gilt nun mit $\mu = (\mu_1, \dots, \mu_n)^\top$, dass

$$\mathbb{E}(R(w)) = \mathbb{E}(w_1 R_1 + \dots + w_n R_n) = w_1 \mu_1 + \dots + w_n \mu_n = \mu^\top w.$$

2. Das „Mean-Variance“-Modell nach Markowitz

Für die Varianz des Portfolio>Returns erhalten wir

$$\text{Var}(R(w)) = \mathbb{E} \left[\left(\sum_{i=1}^n w_i R_i - \mathbb{E} \left(\sum_{i=1}^n w_i R_i \right) \right)^2 \right] = \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[(R_i - \mu_i)(R_j - \mu_j)] w_i w_j.$$

Sei nun $\sigma_{i,j} := \text{Cov}(R_i, R_j) = \mathbb{E}[(R_i - \mu_i)(R_j - \mu_j)]$ für $i \neq j$ und $\sigma_i^2 := \sigma_{i,i} = \text{Var}(R_i) = \mathbb{E}[(R_i - \mu_i)^2]$. Die Kovarianzmatrix der Returns der n Assets lässt sich also schreiben als

$$\Sigma := \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_{n-1,n} \\ \sigma_{n,1} & \cdots & \sigma_{n,n-1} & \sigma_n^2 \end{bmatrix} = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_{2,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_{n-1,n} \\ \sigma_{n,1} & \cdots & \sigma_{n,n-1} & \sigma_{n,n} \end{bmatrix}$$

Mit dieser Notation erhält man

$$\text{Var}(R(w)) = \sum_{i=1}^n \sum_{j=1}^n \sigma_{i,j} w_i w_j = w^\top \Sigma w.$$

Als Kovarianzmatrix ist Σ symmetrisch und positiv semidefinit. Es gilt nämlich für alle $w \in \mathbb{R}^n$, dass $w^\top \Sigma w = \text{Var}(R(w)) \geq 0$, da Varianzen immer nichtnegativ sind.

Des Weiteren können obere und untere Schranken für die Gewichte w_i angenommen werden. Wir betrachten also Restriktionen der Form $L_i \leq w_i \leq U_i$, $i = 1, \dots, n$. Dies schließt auch den Spezialfall ein, in welchem nur „Long Positions“ eingenommen werden dürfen (d.h. $w_i \geq 0$, $\forall i = 1, \dots, n$).

Das Markowitz'sche Modell zielt nun darauf ab, ein Portfolio mit minimaler Varianz, welche Markowitz als Risikomaß verwendet, zu finden, und zwar unter all jenen Portfolios, die einen erwarteten Return in der Höhe von r (oder größer) aufweisen. Die entscheidende Annahme, dass jeder Investor einen hohen erwarteten Return als wünschenswert *und* eine hohe Varianz des Returns als unerwünscht ansieht, bezeichnete Markowitz in seinem ersten Paper „*Portfolio Selection*“ von 1952 als „*expected returns - variance of returns*“-Regel.

Das resultierende „Mean-Variance“-Optimierungsproblem lässt sich folgendermaßen als quadratisches Programm angeben:

$$\begin{aligned} \text{(MW0)} \quad & \min_{w \in \mathbb{R}^n} \quad \frac{1}{2} w^\top \Sigma w \\ & \text{s.t.} \quad \mu^\top w \geq r \\ & \quad \quad e^\top w = 1 \\ & \quad \quad L_i \leq w_i \leq U_i, \quad i = 1, \dots, n \end{aligned}$$

Im Zusammenhang mit diesem Modell verwenden wir die Begriffe „Varianz“ und „Risiko“ eines Portfolios w synonym für die Varianz des Portfolio>Returns $R(w)$.

2. Das „Mean-Variance“-Modell nach Markowitz

Die Zielfunktion dieses Problems ist konvex, da Σ positiv semidefinit ist (siehe Kapitel 1.3). Gelingt es uns also, ein lokales Minimum von (MW0) zu finden, so wissen wir, dass dies bereits ein globales Minimum ist.

Das Optimierungsproblem (MW0) lässt sich nun in folgende (an (QP) angelehnte) Form bringen, mit welcher wir weiterarbeiten möchten:

$$\begin{aligned}
 \text{(MW1)} \quad & \min_{w \in \mathbb{R}^n} \quad \frac{1}{2} w^\top \Sigma w \\
 & \text{s.t.} \quad \mu^\top w \geq r \\
 & \quad \quad Aw \geq \alpha \\
 & \quad \quad Bw = \beta
 \end{aligned}$$

Dabei sind $A \in \mathbb{R}^{p \times n}$, $\alpha \in \mathbb{R}^p$, $B \in \mathbb{R}^{q \times n}$ und $\beta \in \mathbb{R}^q$.

Wir definieren zur Notationsvereinfachung

- $r_{\min} := \min_{w \in \mathbb{R}^n} \{ \mu^\top w \mid Aw \geq \alpha, Bw = \beta \}$,
- $r_{\max} := \max_{w \in \mathbb{R}^n} \{ \mu^\top w \mid Aw \geq \alpha, Bw = \beta \}$,
- $\sigma_{\min}^2 := \min_{w \in \mathbb{R}^n} \{ w^\top \Sigma w \mid Aw \geq \alpha, Bw = \beta \}$,

und nehmen in weiterer Folge an, dass $r \leq r_{\max}$, damit (MW1) überhaupt zulässig ist.

2.2. Die effiziente Front

Wir möchten nun untersuchen, welche Portfolios für Investoren prinzipiell erstrebenswert sind. Diese sogenannten *effizienten* Portfolios lassen sich wie folgt definieren:

Definition. (Effizientes Portfolio)

Ein zulässiges Portfolio $w^* \in \Omega_{MW1}$ mit

$$\Omega_{MW1} := \{ w \in \mathbb{R}^n \mid \mu^\top w \geq r, Aw \geq \alpha \text{ und } Bw = \beta \}$$

heißt **effizient**, falls es die kleinste Varianz unter allen Portfolios besitzt, welche mindestens einen gewissen erwarteten Return aufweisen, oder falls es den größten erwarteten Return unter all jenen Portfolios hat, deren Varianzen ein gewisses Level nicht übersteigen. Anders formuliert müssen für ein effizientes Portfolio $w^* \in \Omega_{MW1}$ die beiden folgenden Bedingungen erfüllt sein:

- $\nexists w \in \Omega_{MW1}$ mit $\mathbb{E}(R(w)) > \mathbb{E}(R(w^*))$ und $\text{Var}(R(w)) \leq \text{Var}(R(w^*))$
- $\nexists w \in \Omega_{MW1}$ mit $\mathbb{E}(R(w)) \geq \mathbb{E}(R(w^*))$ und $\text{Var}(R(w)) < \text{Var}(R(w^*))$

2. Das „Mean-Variance“-Modell nach Markowitz

Effiziente Portfolios sind also Optima des Problems (MW1) und **Pareto-optimal** (oder Pareto-effizient) für folgendes multikriterielle Optimierungsproblem (siehe Appendix A.1):

$$\begin{aligned} \min_{w \in \mathbb{R}^n} \quad & \frac{1}{2} w^\top \Sigma w \\ \max_{w \in \mathbb{R}^n} \quad & \mu^\top w \\ \text{s.t.} \quad & Aw \geq \alpha \\ & Bw = \beta \end{aligned}$$

Mithilfe der Menge aller effizienten Portfolios definieren wir nun die effiziente Front:

Definition. (Effiziente Front)

Sei für ein $r \in [r_{\min}, r_{\max}]$ das Portfolio $w^{(r)}$ ein Minimum von (MW1) mit erwartetem Return $r_{w^{(r)}}^* := \mu^\top w^{(r)} \geq r$ und Portfolio-Varianz $\sigma_r^2 := w^{(r)\top} \Sigma w^{(r)}$. Dann heißt die Menge

$$\left\{ (r_{w^{(r)}}^*, \sigma_r^2) \mid r_{w^{(r)}}^* = \mu^\top w^{(r)}, \sigma_r^2 = w^{(r)\top} \Sigma w^{(r)}, r \in [r_{\min}, r_{\max}] \right\}$$

effiziente Front (efficient frontier) von (MW1). Für den Fall, dass es zu einem r mehrere verschiedene Portfolios mit minimaler Varianz (und eventuell unterschiedlich hohen Returns) gibt, müssen *alle* minimalen Portfolios in obiger Menge berücksichtigt werden.

Bereits 1952 stellte Markowitz grundlegende Überlegungen bezüglich dieser effizienten Front an. Wir möchten nun folgende nützliche Eigenschaft, welche die Konvexität der effizienten Front impliziert, beweisen:

Satz 2.2.1. *Betrachte folgendes Optimierungsproblem, welches (MW1) dahingehend abwandelt, dass der erwartete Return nicht nach unten beschränkt, sondern durch eine Gleichung fixiert wird:*

$$\begin{aligned} (\overline{MW1}) \quad & \min_{w \in \mathbb{R}^n} \quad \frac{1}{2} w^\top \Sigma w \\ \text{s.t.} \quad & \mu^\top w = r \\ & Aw \geq \alpha \\ & Bw = \beta \end{aligned}$$

Sei nun $\bar{w}^{(r)}$ eine Optimallösung von $(\overline{MW1})$ mit Risiko $\bar{\sigma}_r^2 := \bar{w}^{(r)\top} \Sigma \bar{w}^{(r)}$ für ein $r \in [r_{\min}, r_{\max}]$. Dann ist

$$\begin{aligned} \bar{\sigma}_r^2 : [r_{\min}, r_{\max}] &\rightarrow \mathbb{R}_0^+ \\ r &\mapsto \bar{w}^{(r)\top} \Sigma \bar{w}^{(r)} \end{aligned} \tag{2.1}$$

eine konvexe Funktion. Falls Σ positiv definit und damit die Zielfunktion von $(\overline{MW1})$ strikt konvex ist, so ist auch die obige Abbildung strikt konvex.

2. Das „Mean-Variance“-Modell nach Markowitz

Beweis. (Vgl. Steinbach, 2001.)

Seien $r_{\min} \leq r_1 \leq r_{\max}$ sowie $r_{\min} \leq r_2 \leq r_{\max}$ mit $r_1 < r_2$ und $r_3 := \lambda r_1 + (1 - \lambda)r_2$ mit $\lambda \in (0, 1)$ beliebig.

Weiters seien $\bar{w}^{(r_1)}, \bar{w}^{(r_2)}$ und $\bar{w}^{(r_3)}$ die Optimallösungen von $(\overline{\text{MW1}})$, wenn r auf den Wert r_1, r_2 bzw. r_3 gesetzt wird.

$\bar{w} := \lambda \bar{w}^{(r_1)} + (1 - \lambda) \bar{w}^{(r_2)}$ ist zulässig, da $\bar{w}^{(r_1)}$ und $\bar{w}^{(r_2)}$ zulässig sind und die Menge der zulässigen Lösungen konvex ist. Wir erhalten nun

$$\mu^\top \bar{w} = \lambda \mu^\top \bar{w}^{(r_1)} + (1 - \lambda) \mu^\top \bar{w}^{(r_2)} = \lambda r_1 + (1 - \lambda)r_2 = r_3$$

für den erwarteten Return des Portfolios \bar{w} . Aus der Konvexität der Zielfunktion von $(\overline{\text{MW1}})$ folgt

$$\bar{\sigma}_{r_3}^2 = \bar{w}^{(r_3)\top} \Sigma \bar{w}^{(r_3)} \leq \bar{w}^\top \Sigma \bar{w} \leq \lambda \bar{w}^{(r_1)\top} \Sigma \bar{w}^{(r_1)} + (1 - \lambda) \bar{w}^{(r_2)\top} \Sigma \bar{w}^{(r_2)} = \lambda \bar{\sigma}_{r_1}^2 + (1 - \lambda) \bar{\sigma}_{r_2}^2,$$

wobei im Falle positiver Definitheit der Kovarianzmatrix die letzte Zeile eine strikte Ungleichung ist. Damit ist die Funktion in (2.1) konvex (bzw. strikt konvex). \square

Wir können nun mithilfe des soeben bewiesenen Satzes die effiziente Front von (MW1) folgendermaßen charakterisieren:

Seien $w^{(r)}, r_{w^{(r)}}^*, \sigma_r^2$ für (MW1) und $\bar{w}^{(r)}, \bar{\sigma}_r^2$ für $(\overline{\text{MW1}})$ mit $r \in [r_{\min}, r_{\max}]$ wie oben. Dann gilt offensichtlich

$$\sigma_r^2 = \min\{\bar{\sigma}_{r'}^2 \mid r' \in [r, r_{\max}]\}.$$

Für die effiziente Front gilt

$$\begin{aligned} & \left\{ (r_{w^{(r)}}^*, \sigma_r^2) \mid r_{w^{(r)}}^* = \mu^\top w^{(r)}, \sigma_r^2 = w^{(r)\top} \Sigma w^{(r)}, r \in [r_{\min}, r_{\max}] \right\} \\ & = \left\{ (r, \bar{\sigma}_r^2) \mid \bar{\sigma}_r^2 = \bar{w}^{(r)\top} \Sigma \bar{w}^{(r)}, r \in [\hat{r}, r_{\max}] \right\}, \end{aligned}$$

wobei \hat{r} jener Wert ist, für den man $\bar{\sigma}_{\hat{r}}^2 = \sigma_{\min}^2$ erhält.

Unter der vereinfachenden Annahme, dass $\Sigma \succ 0$, ist \hat{r} natürlich eindeutig, da hier die Zielfunktion von $(\overline{\text{MW1}})$ laut Annahme strikt konvex ist. Betrachten wir nämlich $(\overline{\text{MW1}})$ ohne die „Return-Restriktion“ $\mu^\top w = r$, so existiert ein eindeutiges Portfolio als Minimum (also mit Zielfunktionswert $\frac{1}{2} \sigma_{\min}^2$) und der Return dieses Portfolios entspricht genau \hat{r} .

Somit wird die effiziente Front des Problems (MW1) durch die Funktion in (2.1) mit eingeschränktem Definitionsbereich beschrieben und ist aufgrund des vorherigen Satzes strikt konvex. Offenbar erfüllt eine Optimallösung von (MW1) die Ungleichheitsrestriktion $\mu^\top w \geq r$ genau dann strikt, wenn das Problem für $r \in [r_{\min}, \hat{r}]$ gelöst wird.

Mit $\Sigma \succcurlyeq 0$, jedoch ohne die Bedingung $\Sigma \succ 0$, ist die effiziente Front konvex, aber nicht zwangsweise strikt konvex. Es ist zu beachten, dass dann in (MW1) zu einem r im Allgemeinen verschiedene Optima $w_1^{(r)}$ und $w_2^{(r)}$ mit $r_{w_1^{(r)}}^* \neq r_{w_2^{(r)}}^*$ existieren können (vgl.

2. Das „Mean-Variance“-Modell nach Markowitz

unsere Definition der effizienten Front). In diesem Fall muss für \hat{r} der kleinste unter all jenen Werten gewählt werden, die zum Zielfunktionswert $\frac{1}{2}\sigma_{\min}^2$ in (MW1) führen, also

$$\hat{r} := \min\{r \in [r_{\min}, r_{\max}] \mid \bar{\sigma}_r^2 = \sigma_{\min}^2\}.$$

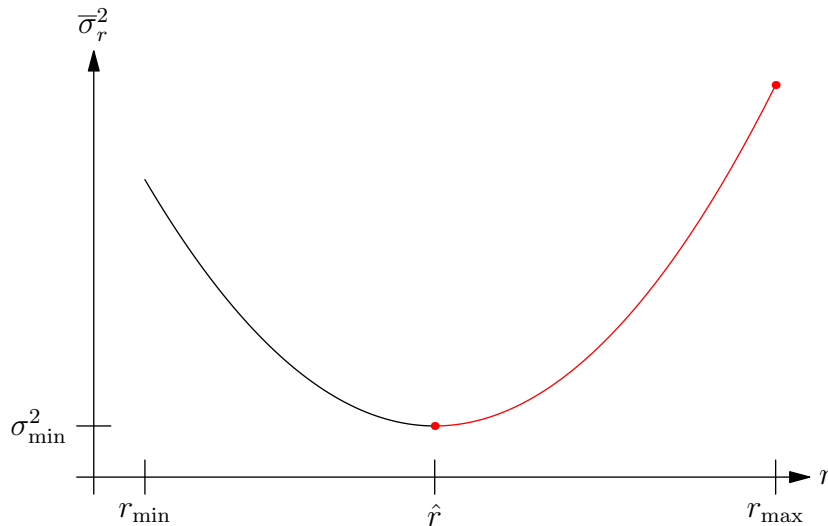


Abb. 2.1.: Die effiziente Front des „Mean-Variance“-Modells (in rot).

2.3. KKT-Bedingungen und äquivalente Modellformulierungen

Als nächstes geben wir die KKT-Bedingungen von (MW1) an:

Korollar 2.3.1. *Ein Punkt $w^{(r)} \in \mathbb{R}^n$ ist (für ein fixes $r \leq r_{\max}$) genau dann ein globales Minimum von (MW1), falls $\lambda^{(r)} \in \mathbb{R}$, $\lambda^{(1)} \in \mathbb{R}^p$ und $\gamma^{(1)} \in \mathbb{R}^q$ existieren, sodass folgende KKT-Bedingungen erfüllt sind:*

- [1-i] $\mu^\top w^{(r)} \geq r$, $Aw^{(r)} \geq \alpha$, $Bw^{(r)} = \beta$, d.h. $w^{(r)}$ ist zulässig für (MW1),
- [1-ii] $\Sigma w^{(r)} - \lambda^{(r)}\mu - A^\top \lambda^{(1)} - B^\top \gamma^{(1)} = 0$,
- [1-iii] $\lambda^{(r)} \geq 0$, $\lambda^{(r)} \cdot (\mu^\top w^{(r)} - r) = 0$,
- [1-iv] $\lambda^{(1)} \geq 0$, $\lambda^{(1)\top} \cdot (Aw^{(r)} - \alpha) = 0$.

Beweis. Folgt direkt aus Korollar 1.4.8. □

2. Das „Mean-Variance“-Modell nach Markowitz

Wir möchten nun noch zwei äquivalente Darstellungen von (MW1) genauer untersuchen:

$$\begin{array}{ll}
 \text{(MW2)} & \max_{w \in \mathbb{R}^n} \mu^\top w \\
 & \text{s.t. } w^\top \Sigma w \leq \sigma^2 \\
 & \quad Aw \geq \alpha \\
 & \quad Bw = \beta \\
 \text{(MW3)} & \max_{w \in \mathbb{R}^n} (1 - \delta) \cdot \mu^\top w - \delta \cdot \left(\frac{1}{2} w^\top \Sigma w \right) \\
 & \text{s.t. } Aw \geq \alpha \\
 & \quad Bw = \beta
 \end{array}$$

Dabei sei $\delta \in [0, 1]$ ein Gewichtsparameter, welcher den Trade-off zwischen dem erwarteten Return und dem Risiko (Varianz) des Portfolios steuert. Während verschiedene Werte von r in (MW1) bzw. σ^2 in (MW2) - welche sich auf die *Restriktionen* der entsprechenden Probleme auswirken - verschiedene effiziente Portfolios auf der effizienten Front erzeugen, geschieht dies bei (MW3) durch Variation des Parameters δ , der die *Zielfunktion* des Problems beeinflusst.

Wir bringen diese beiden Probleme nun noch in unsere kanonische Form (NLP) und berechnen, wie bereits für (MW1), die KKT-Bedingungen, welche wir für den Beweis der Äquivalenz dieser drei Varianten des „Mean-Variance“-Modells benötigen werden:

$$\begin{array}{ll}
 \text{(MW2)} & \min_{w \in \mathbb{R}^n} -\mu^\top w \\
 & \text{s.t. } -w^\top \Sigma w \geq -\sigma^2 \\
 & \quad Aw \geq \alpha \\
 & \quad Bw = \beta \\
 \text{(MW3)} & \min_{w \in \mathbb{R}^n} \delta \cdot \left(\frac{1}{2} w^\top \Sigma w \right) - (1 - \delta) \cdot \mu^\top w \\
 & \text{s.t. } Aw \geq \alpha \\
 & \quad Bw = \beta
 \end{array}$$

Problem (MW2) ist nun (nach unserer Definition) kein quadratisches Problem (QP) mehr, da hier auch nichtlineare Ungleichheitsrestriktionen auftauchen. Die „problematische“ Restriktion $-w^\top \Sigma w \geq -\sigma^2 \Leftrightarrow \sigma^2 - w^\top \Sigma w \geq 0$ ist allerdings aufgrund der positiven Semidefinitheit von Σ konkav, womit (aufgrund der linearen und somit auch konvexen Zielfunktion) ein konvexes Programm (KP) vorliegt. Wir können daher folgende Aussage formulieren:

Korollar 2.3.2. *Falls für einen Punkt $w^{(\sigma^2)} \in \mathbb{R}^n$ (für ein fixes $\sigma^2 \geq \sigma_{\min}^2$) die Vektoren $\lambda^{(\sigma^2)} \in \mathbb{R}$, $\lambda^{(2)} \in \mathbb{R}^p$ und $\gamma^{(2)} \in \mathbb{R}^q$ existieren, sodass die KKT-Bedingungen*

$$\begin{array}{l}
 [2-i] \quad -w^{(\sigma^2)\top} \Sigma w^{(\sigma^2)} \geq -\sigma^2, \quad Aw^{(\sigma^2)} \geq \alpha, \quad Bw^{(\sigma^2)} = \beta, \\
 [2-ii] \quad -\mu + \lambda^{(\sigma^2)} \Sigma w^{(\sigma^2)} - A^\top \lambda^{(2)} - B^\top \gamma^{(2)} = 0, \\
 [2-iii] \quad \lambda^{(\sigma^2)} \geq 0, \quad \lambda^{(\sigma^2)} \cdot (\sigma^2 - w^{(\sigma^2)\top} \Sigma w^{(\sigma^2)}) = 0, \\
 [2-iv] \quad \lambda^{(2)} \geq 0, \quad \lambda^{(2)\top} \cdot (Aw^{(\sigma^2)} - \alpha) = 0,
 \end{array}$$

erfüllt sind, so ist $w^{(\sigma^2)}$ ein globales Minimum von (MW2).

Ist umgekehrt $w^{(\sigma^2)}$ ein (globales) Minimum von (MW2) und erfüllt (MW2) die Regularitätsbedingung von Slater, dann existieren $\lambda^{(\sigma^2)} \in \mathbb{R}$, $\lambda^{(2)} \in \mathbb{R}^p$ und $\gamma^{(2)} \in \mathbb{R}^q$, sodass die obigen KKT-Bedingungen erfüllt sind.

2. Das „Mean-Variance“-Modell nach Markowitz

Beweis.

„ \Rightarrow “: Anwendung von Satz 1.4.7.

„ \Leftarrow “: Anwendung von Satz 1.4.6. □

In diesem Fall ist die Regularitätsbedingung von Slater jedoch nicht besonders einschränkend, da nur eine einzige der Ungleichheitsrestriktionen nichtlinear ist. Für die Erfüllung der Slater-Regularität muss also ein zulässiger Punkt \hat{w} für (MW2) mit $-\hat{w}^\top \Sigma \hat{w} > -\sigma^2$ existieren. Dies ist nur dann verletzt, wenn σ^2 auf einen Wert kleiner oder gleich der minimalen Varianz σ_{\min}^2 gesetzt wird. Für $\sigma^2 > \sigma_{\min}^2$ existiert nämlich (per Definition von σ_{\min}^2) immer ein zulässiges Portfolio, welches minimale Varianz σ_{\min}^2 besitzt und damit klarerweise auch ein Punkt ist, welcher die Bedingungen der Slater-Regularität erfüllt. Da für $\sigma^2 < \sigma_{\min}^2$ das Problem (MW2) ohnehin unzulässig ist, gehen wir davon aus, dass stets $\sigma^2 \geq \sigma_{\min}^2$.

Hingegen entspricht (MW3) wieder einem klassischen quadratischen Programm mit konvexer Zielfunktion und ausschließlich linearen Restriktionen. Setzen wir für den Parameter δ Werte im Intervall $[0, 1]$ voraus, so beinhaltet dies auch zwei Spezialfälle: Mit $\delta = 0$ erhalten wir eine Maximierung des erwarteten Return (unabhängig vom Risiko des Portfolios), während $\delta = 1$ zu einer Minimierung des Risikos (unabhängig vom erwarteten Return) führt. Analog zu (MW1) gilt:

Korollar 2.3.3. *Ein Punkt $w^{(\delta)} \in \mathbb{R}^n$ ist (für ein fixes $\delta \in [0, 1]$) genau dann ein globales Minimum von (MW3), falls $\lambda^{(3)} \in \mathbb{R}^p$ und $\gamma^{(3)} \in \mathbb{R}^q$ existieren, sodass folgende KKT-Bedingungen erfüllt sind:*

$$\begin{aligned} [\beta\text{-i}] \quad & Aw^{(\delta)} \geq \alpha, \quad Bw^{(\delta)} = \beta, \\ [\beta\text{-ii}] \quad & \delta \Sigma w^{(\delta)} - (1 - \delta)\mu - A^\top \lambda^{(3)} - B^\top \gamma^{(3)} = 0, \\ [\beta\text{-iii}] \quad & \lambda^{(3)} \geq 0, \quad \lambda^{(3)\top} \cdot (Aw^{(\delta)} - \alpha) = 0. \end{aligned}$$

Beweis. Folgt direkt aus Korollar 1.4.8. □

Folgender Satz zeigt uns nun die Äquivalenz aller drei Optimierungsprobleme:

Satz 2.3.4. *(Äquivalenz von (MW1), (MW2) und (MW3))*

- Äquivalenz von (MW1) und (MW3):
 - a) Sei $w^{(r)} \in \mathbb{R}^n$ ein Minimum von (MW1) für ein $r \leq r_{\max}$. Dann existiert ein $\delta \in (0, 1]$, sodass $w^{(r)}$ auch ein Minimum von (MW3) ist.
 - b) Sei $w^{(\delta)} \in \mathbb{R}^n$ ein Minimum von (MW3) für ein $\delta \in (0, 1]$. Dann existieren ein $r \leq r_{\max}$, sodass $w^{(\delta)}$ auch ein Minimum von (MW1) ist.
- Äquivalenz von (MW2) und (MW3):
 - c) Sei $w^{(\sigma^2)} \in \mathbb{R}^n$ ein Minimum von (MW2) für ein $\sigma^2 > \sigma_{\min}^2$. Dann existiert ein $\delta \in [0, 1)$, sodass $w^{(\sigma^2)}$ auch ein Minimum von (MW3) ist.

2. Das „Mean-Variance“-Modell nach Markowitz

- d) Sei $w^{(\sigma_{\min}^2)} \in \mathbb{R}^n$ ein Minimum von (MW2) für $\sigma^2 = \sigma_{\min}^2$. Dann ist $w^{(\sigma^2)}$ auch ein Minimum von (MW3) mit $\delta = 1$.
- e) Sei $w^{(\delta)} \in \mathbb{R}^n$ ein Minimum von (MW3) für ein $\delta \in [0, 1)$. Dann existiert ein $\sigma^2 \geq \sigma_{\min}^2$, sodass $w^{(\delta)}$ auch ein Minimum von (MW2) ist.

Beweis. Wir beweisen die Äquivalenz der drei Optimierungsprobleme unter Verwendung der Korollare 2.3.1, 2.3.2 und 2.3.3 mit den entsprechenden KKT-Bedingungen für (MW1), (MW2) und (MW3).

- a) Betrachte den KKT-Punkt $(w^{(r)}, \lambda^{(r)}, \lambda^{(1)}, \gamma^{(1)})$, welcher laut Korollar 2.3.1 existiert. Dann ist der zugehörige Lagrange-Multiplikator $\lambda^{(r)} \geq 0$ (wegen [1-iii]). Damit lautet Bedingung [1-ii]:

$$\begin{aligned} \Sigma w^{(r)} - \lambda^{(r)} \mu - A^\top \lambda^{(1)} - B^\top \gamma^{(1)} &= 0 & \Leftrightarrow \\ \frac{1}{1 + \lambda^{(r)}} \Sigma w^{(r)} - \frac{\lambda^{(r)}}{1 + \lambda^{(r)}} \mu - A^\top \frac{\lambda^{(1)}}{1 + \lambda^{(r)}} - B^\top \frac{\gamma^{(1)}}{1 + \lambda^{(r)}} &= 0 \end{aligned}$$

Für Bedingung [1-iv] erhalten wir:

$$\lambda^{(1)} \geq 0, \lambda^{(1)\top} \cdot (Aw^{(r)} - \alpha) = 0 \quad \Leftrightarrow \quad \frac{\lambda^{(1)}}{1 + \lambda^{(r)}} \geq 0, \frac{\lambda^{(1)\top}}{1 + \lambda^{(r)}} \cdot (Aw^{(r)} - \alpha) = 0$$

Mit der Wahl

$$\begin{aligned} \lambda^{(3)} &:= \frac{\lambda^{(1)}}{1 + \lambda^{(r)}} \geq 0, \\ \gamma^{(3)} &:= \frac{\gamma^{(1)}}{1 + \lambda^{(r)}}, \\ \delta &:= \frac{1}{1 + \lambda^{(r)}} \in (0, 1] \Rightarrow \frac{\lambda^{(r)}}{1 + \lambda^{(r)}} = 1 - \delta \in [0, 1), \end{aligned}$$

erfüllt daher der Punkt $(w^{(r)}, \lambda^{(3)}, \gamma^{(3)})$ die KKT-Bedingungen [3-i], [3-ii] sowie [3-iii] und ist damit ein Minimum von (MW3).

- b) Betrachte den KKT-Punkt $(w^{(\delta)}, \lambda^{(3)}, \gamma^{(3)})$, welcher laut Korollar 2.3.3 existiert. Unter der Annahme, dass $\delta \in (0, 1]$, können wir Bedingung [3-ii] folgendermaßen umformen:

$$\begin{aligned} \delta \Sigma w^{(\delta)} - (1 - \delta) \mu - A^\top \lambda^{(3)} - B^\top \gamma^{(3)} &= 0 & \Leftrightarrow \\ \Sigma w^{(\delta)} - \frac{1 - \delta}{\delta} \mu - A^\top \frac{\lambda^{(3)}}{\delta} - B^\top \frac{\gamma^{(3)}}{\delta} &= 0 \end{aligned}$$

Für Bedingung [3-iii] erhalten wir:

$$\lambda^{(3)} \geq 0, \lambda^{(3)\top} \cdot (Aw^{(\delta)} - \alpha) = 0 \quad \Leftrightarrow \quad \frac{\lambda^{(3)}}{\delta} \geq 0, \frac{\lambda^{(3)\top}}{\delta} \cdot (Aw^{(\delta)} - \alpha) = 0$$

2. Das „Mean-Variance“-Modell nach Markowitz

Mit der Wahl

$$\begin{aligned}\lambda^{(r)} &:= \frac{1 - \delta}{\delta} \geq 0, \\ \lambda^{(1)} &:= \frac{\lambda^{(3)}}{\delta} \geq 0, \\ \gamma^{(1)} &:= \frac{\gamma^{(3)}}{\delta}, \\ r &:= \mu^\top w^{(\delta)} \leq r_{\max} \Rightarrow \mu^\top w^{(\delta)} - r = 0,\end{aligned}$$

erfüllt daher der Punkt $(w^{(\delta)}, \lambda^{(r)}, \lambda^{(1)}, \gamma^{(1)})$ die KKT-Bedingungen [1-*i*] bis [1-*iv*] und ist damit ein Minimum von (MW1).

- c) Wegen $\sigma^2 > \sigma_{\min}^2$ ist die Regularitätsbedingung von Slater erfüllt und es existiert laut Korollar 2.3.2 ein KKT-Punkt $(w^{(\sigma^2)}, \lambda^{(\sigma^2)}, \lambda^{(2)}, \gamma^{(2)})$.

Es gilt, dass der zugehörige Lagrange-Multiplikator $\lambda^{(\sigma^2)} \geq 0$ ist (wegen [2-*iii*]). Damit lautet Bedingung [2-*ii*]:

$$\begin{aligned}-\mu + \lambda^{(\sigma^2)} \Sigma w^{(\sigma^2)} - A^\top \lambda^{(2)} - B^\top \gamma^{(2)} &= 0 \quad \Leftrightarrow \\ -\frac{1}{1 + \lambda^{(\sigma^2)}} \mu + \frac{\lambda^{(\sigma^2)}}{1 + \lambda^{(\sigma^2)}} \Sigma w^{(\sigma^2)} - A^\top \frac{\lambda^{(2)}}{1 + \lambda^{(\sigma^2)}} - B^\top \frac{\gamma^{(2)}}{1 + \lambda^{(\sigma^2)}} &= 0\end{aligned}$$

Für Bedingung [2-*iv*] erhalten wir:

$$\lambda^{(2)} \geq 0, \lambda^{(2)\top} \cdot (Aw^{(\sigma^2)} - \alpha) = 0 \quad \Leftrightarrow \quad \frac{\lambda^{(2)}}{1 + \lambda^{(\sigma^2)}} \geq 0, \frac{\lambda^{(2)\top}}{1 + \lambda^{(\sigma^2)}} \cdot (Aw^{(\sigma^2)} - \alpha) = 0$$

Mit der Wahl

$$\begin{aligned}\lambda^{(3)} &:= \frac{\lambda^{(2)}}{1 + \lambda^{(\sigma^2)}} \geq 0, \\ \gamma^{(3)} &:= \frac{\gamma^{(2)}}{1 + \lambda^{(\sigma^2)}}, \\ \delta &:= \frac{\lambda^{(\sigma^2)}}{1 + \lambda^{(\sigma^2)}} \in [0, 1) \Rightarrow \frac{1}{1 + \lambda^{(\sigma^2)}} = 1 - \delta \in (0, 1],\end{aligned}$$

erfüllt daher der Punkt $(w^{(\sigma^2)}, \lambda^{(3)}, \gamma^{(3)})$ die KKT-Bedingungen [3-*i*], [3-*ii*] sowie [3-*iii*] und ist damit ein Minimum von (MW3).

- d) Mit $\sigma^2 = \sigma_{\min}^2$ und $\delta = 1$ erhalten wir folgende Optimierungsprobleme:

$$\begin{array}{ll} \widehat{\text{(MW2)}} & \min_{w \in \mathbb{R}^n} -\mu^\top w \\ & \text{s.t. } w^\top \Sigma w = \sigma_{\min}^2 \\ & Aw \geq \alpha \\ & Bw = \beta \end{array} \qquad \begin{array}{ll} \widehat{\text{(MW3)}} & \min_{w \in \mathbb{R}^n} \frac{1}{2} w^\top \Sigma w \\ & \text{s.t. } Aw \geq \alpha \\ & Bw = \beta \end{array}$$

2. Das „Mean-Variance“-Modell nach Markowitz

Wir sehen, dass die Menge der Minima in $(\widetilde{\text{MW3}})$ genau der Menge aller zulässigen Portfolios des Problems $(\widetilde{\text{MW2}})$ entspricht. Damit folgt das zu Beweisende.

- e) Betrachte den KKT-Punkt $(w^{(\delta)}, \lambda^{(3)}, \gamma^{(3)})$, welcher laut Korollar 2.3.3 existiert. Unter der Annahme, dass $\delta \in [0, 1)$, können wir Bedingung [3-ii] folgendermaßen umformen:

$$\begin{aligned} \delta \Sigma w^{(\delta)} - (1 - \delta) \mu - A^\top \lambda^{(3)} - B^\top \gamma^{(3)} &= 0 & \Leftrightarrow \\ \frac{\delta}{1 - \delta} \Sigma w^{(\delta)} - \mu - A^\top \frac{\lambda^{(3)}}{1 - \delta} - B^\top \frac{\gamma^{(3)}}{1 - \delta} &= 0 \end{aligned}$$

Für Bedingung [3-iii] erhalten wir:

$$\lambda^{(3)} \geq 0, \lambda^{(3)\top} \cdot (Aw^{(\delta)} - \alpha) = 0 \quad \Leftrightarrow \quad \frac{\lambda^{(3)}}{1 - \delta} \geq 0, \frac{\lambda^{(3)\top}}{1 - \delta} \cdot (Aw^{(\delta)} - \alpha) = 0$$

Mit der Wahl

$$\begin{aligned} \lambda^{(\sigma^2)} &:= \frac{\delta}{1 - \delta} \geq 0, \\ \lambda^{(2)} &:= \frac{\lambda^{(3)}}{1 - \delta} \geq 0, \\ \gamma^{(2)} &:= \frac{\gamma^{(3)}}{1 - \delta}, \\ \sigma^2 &:= w^{(\delta)\top} \Sigma w^{(\delta)} \geq \sigma_{\min}^2 \Rightarrow w^{(\delta)\top} \Sigma w^{(\delta)} - \sigma^2 = 0, \end{aligned}$$

erfüllt daher der Punkt $(w^{(\delta)}, \lambda^{(\sigma^2)}, \lambda^{(2)}, \gamma^{(2)})$ die KKT-Bedingungen [2-i] bis [2-iv] und ist damit ein Minimum von (MW2) .

□

Es seien an dieser Stelle noch zwei interessante Beobachtungen bezüglich der Äquivalenz der drei angegebenen Probleme erwähnt:

- Wir analysieren den Zusammenhang zwischen (MW1) und (MW3) für den Fall $r = r_{\max}$ und $\delta = 0$: Mit dieser Parameterwahl erhalten wir folgende Probleme:

$$\begin{array}{ll} (\widetilde{\text{MW1}}) & \min_{w \in \mathbb{R}^n} \frac{1}{2} w^\top \Sigma w & (\widetilde{\text{MW3}}) & \min_{w \in \mathbb{R}^n} -\mu^\top w \\ & \text{s.t. } \mu^\top w = r_{\max} & & \text{s.t. } Aw \geq \alpha \\ & & & Bw = \beta \\ & & & Bw = \beta \end{array}$$

Eine Optimallösung von $(\widetilde{\text{MW1}})$ ist zwar offensichtlich auch minimal für $(\widetilde{\text{MW3}})$, die Umkehrung gilt jedoch im Allgemeinen nicht. Dies liegt daran, dass $(\widetilde{\text{MW3}})$

2. Das „Mean-Variance“-Modell nach Markowitz

keine *strikt* konvexe Zielfunktion hat, weshalb es nicht ausgeschlossen ist, dass dieses Optimierungsproblem mehrere globale Minima hat.

Zu beachten ist jedoch, dass ein Minimum von $(\widehat{MW1})$ nach Satz 2.3.4 *a)* auch $(MW3)$ mit einem $\delta > 0$ minimiert, wobei δ klarerweise hinreichend klein sein muss: Wir möchten ja ein Portfolio mit maximalem Return finden und die Varianz nur insofern berücksichtigen, dass wir unter allen Portfolios mit größtmöglichem Return jenes mit dem kleinsten Risiko erhalten.

- Analog zu vorheriger Beobachtung betrachten wir nochmals Satz 2.3.4 *d)*. Die Umkehrung der Aussage dieses Satzes gilt, falls Σ positiv definit und damit die Zielfunktion von $(\widehat{MW3})$ strikt konvex ist. Unter dieser Annahme ist nämlich ein Minimum $w^* \in \mathbb{R}^n$ dieses Problems das eindeutige, globale Minimum. Aus diesem Grund ist w^* der einzige zulässige Punkt für $(\widehat{MW2})$ und damit klarerweise auch dessen eindeutige Optimallösung. Ohne die Voraussetzung $\Sigma \succ 0$ herrscht jedoch in 2.3.4 *d)* im Allgemeinen keine Äquivalenz: $(\widehat{MW3})$ könnte dann mehrere Minima mit unterschiedlich hohen Returns haben.

Satz 2.3.4 *c)* liefert uns auch keine Information darüber, ob ein Minimum von $(\widehat{MW2})$ auch ein Minimum von $(MW3)$ mit einem $\delta < 1$ ist, da ja $(\widehat{MW2})$ die Slater-Regularität nicht erfüllt.

3. Risikobasierte Portfoliooptimierung

Im Rahmen der risikobasierten Portfoliooptimierung werden wir nun folgende Herangehensweisen zur Bestimmung von Assetportfolios kennenlernen:

- Den „Minimum-Variance“-Ansatz
- Den „Risk-Budgeting“- und „Risk-Parity“-Ansatz
- Den „Weight-Budgeting“-Ansatz

Einen wichtigen Teil dieses Kapitels werden nichtlineare Optimierungsprobleme darstellen, mithilfe derer die Berechnung sogenannter „Risk-Budgeting“-Portfolios unter bestimmten Bedingungen möglich ist. Für sämtliche auftretende nichtlineare Programme werden die in Abschnitt 1.4 definierten KKT-Bedingungen eine entscheidende Rolle für unsere Untersuchungen spielen.

Als Basis für die Überlegungen der ersten beiden einführenden Abschnitte 3.1 und 3.2 dienen die Papers von Maillard, Roncalli und Teiletche (2010) und Bai, Scheinberg und Tütüncü (2013) sowie das im Jahr 2013 erschienene Buch „Introduction to Risk Parity and Budgeting“ von Thierry Roncalli.

Wie in Kapitel 2 betrachten wir n Assets mit zufälligem Return R_i für das i -te Asset und Kovarianzmatrix

$$\Sigma := \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_{n-1,n} \\ \sigma_{n,1} & \cdots & \sigma_{n,n-1} & \sigma_n^2 \end{bmatrix} = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_{2,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_{n-1,n} \\ \sigma_{n,1} & \cdots & \sigma_{n,n-1} & \sigma_{n,n} \end{bmatrix}$$

wobei wieder $\sigma_{i,j} := \text{Cov}(R_i, R_j)$ für $i \neq j$ und $\sigma_i^2 := \sigma_{i,i} = \text{Var}(R_i)$ sind.

Weiters werden wir $\varrho_{i,j} := \frac{\sigma_{i,j}}{\sigma_i \sigma_j}$ für die Korrelationen der Returns verwenden. Für die Varianz des zufälligen Returns $R(w) := w_1 R_1 + \dots + w_n R_n$ eines Portfolios $w \in \mathbb{R}^n$, welche wir (wie schon beim Markowitz’schen „Mean-Variance“-Modell) als Risikomaß verwenden, erhalten wir also auch die Darstellung

$$\text{Var}(R(w)) = w^\top \Sigma w = \sum_{i=1}^n \sum_{j=1}^n \varrho_{i,j} \sigma_i \sigma_j w_i w_j.$$

3. Risikobasierte Portfoliooptimierung

Wir verwenden in den folgenden Abschnitten die Notation

$$\varsigma^2(w) := \frac{1}{2} w^\top \Sigma w = \frac{1}{2} \left(\sum_{i=1}^n w_i^2 \sigma_i^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n w_i w_j \sigma_{i,j} \right).$$

3.1. Der „Minimum-Variance“-Ansatz

Eine einfache Möglichkeit für ein rein risikobasiertes Portfoliooptimierungsmodell ist die Minimierung der Portfolio-Varianz ohne Berücksichtigung des (erwarteten) Returns. Wir möchten diesen Ansatz als kurze Einführung in das Konzept der risikobasierten Portfoliooptimierung behandeln.

Eine allgemeine Formulierung des „Minimum-Variance“-Optimierungsproblems erhalten wir, wenn wir im quadratischen Programm (MW3), welches wir im 2. Kapitel definiert haben, den Gewichtsparameter δ auf 1 setzen.

Wir möchten nun folgendes Varianzminimierungsmodell mit einer etwas einfacher zu handhabenden Menge der zulässigen Punkte betrachten:

$$\begin{aligned} \text{(MV)} \quad & \min_{w \in \mathbb{R}^n} \quad \frac{1}{2} w^\top \Sigma w \\ & \text{s.t.} \quad w_i \geq 0, \quad i = 1, \dots, n \\ & \quad \quad \sum_{i=1}^n w_i = 1 \end{aligned}$$

Wir beschränken uns also auf „Long-Only“-Portfolios. In (MW3) entspricht dies der Koeffizientenwahl

$$\begin{aligned} A &= I_n \in \mathbb{R}^{n \times n} \text{ (die } n\text{-dimensionale Einheitsmatrix),} \\ \alpha &= 0 \in \mathbb{R}^n, \\ B &= e^\top = (1, \dots, 1) \in \mathbb{R}^{1 \times n}, \\ b &= 1 \in \mathbb{R}^1, \end{aligned}$$

und $\mathbb{R}_{I_n, 1}^n := \{w \in \mathbb{R}^n \mid I_n w \geq 0, e^\top w = 1\}$ als die Menge der zulässigen Portfolios.

Ein KKT-Punkt $(w^*, \lambda, \gamma) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^1$ dieses Problems muss also folgende Bedingungen erfüllen:

$$\begin{aligned} [MV-i] \quad & w^* \geq 0, \quad e^\top w^* = 1, \\ [MV-ii] \quad & \Sigma w^* - \lambda - \gamma e = 0, \\ [MV-iii] \quad & \lambda \geq 0, \quad \lambda^\top w^* = 0. \end{aligned}$$

3. Risikobasierte Portfoliooptimierung

Die Bedingungen des komplementären Schlupfs in $[MV-iii]$ führen gemeinsam mit $[MV-ii]$ und der Bedingung $w_i^* \geq 0, \forall i = 1, \dots, n$ aus $[MV-i]$ zu folgender Forderung:

$$\begin{aligned} (\Sigma w^*)_i &= \gamma, \forall i = 1, \dots, n \text{ mit } w_i^* > 0, \\ &\Leftrightarrow \\ (\Sigma w^*)_i &= (\Sigma w^*)_j, \forall i, j \text{ mit } w_i^*, w_j^* > 0. \end{aligned} \quad (3.1)$$

Wir definieren nun den marginalen Risikobeitrag eines Assets wie folgt:

Definition. (Marginaler Risikobeitrag, vgl. Bai, Scheinberg und Tütüncü (2013))
Betrachte eine Menge von n Assets mit Kovarianzmatrix Σ und eine Menge Ω zulässiger Portfolios. Dann ist der **marginale Risikobeitrag** des i -ten Assets für ein Portfolio $w \in \Omega$ (mit Portfoliorisiko $w^\top \Sigma w = 2 \cdot \zeta^2(w)$) definiert als

$$\partial_{w_i} \zeta^2(w) = \frac{\partial \zeta^2(w)}{\partial w_i} = (\Sigma w)_i = w_i \sigma_i^2 + \sum_{\substack{j=1 \\ j \neq i}}^n w_j \sigma_{i,j}.$$

Damit entspricht die Gleichung (3.1) der Bedingung

$$\left. \frac{\partial \zeta^2(w)}{\partial w_i} \right|_{w=w^*} = \left. \frac{\partial \zeta^2(w)}{\partial w_j} \right|_{w=w^*}, \forall i, j \text{ mit } w_i^*, w_j^* > 0.$$

Der „Minimum-Variance“-Ansatz führt somit zu optimalen Portfolios, für die jedes Asset, in das investiert wird, den gleichen *marginalen* Risikobeitrag besitzt.

3.1.1. Explizite Lösung des „Minimum-Variance“-Problems

Wir interessieren uns nun für eine Lösung des Optimierungsproblems

$$(\overline{MV}) \quad \min_{w \in \mathbb{R}^n} \left\{ \frac{1}{2} w^\top \Sigma w \mid e^\top w = 1 \right\},$$

welches auch Leerverkäufe zulässt und keine Ungleichheitsrestriktionen beinhaltet. Dabei halten wir uns an Maillard, Roncalli und Teiletche (2010) sowie Roncalli (2013).

Wir nehmen an, dass die Matrix Σ positiv definit (und damit invertierbar) ist. Aus der resultierenden strikten Konvexität der Zielfunktion folgt die Eindeutigkeit einer Lösung, welche wir mithilfe der (hinreichenden) KKT-Bedingungen 1. Ordnung in geschlossener Form angeben können:

Die Bedingung $[MV-ii]$ entspricht in diesem Fall der Forderung $w^* = \gamma \cdot \Sigma^{-1} e$. Die Normierungsbedingung $e^\top w^* = 1$ lautet somit

$$\gamma \cdot e^\top \Sigma^{-1} e = 1 \quad \Leftrightarrow \quad \gamma = \frac{1}{e^\top \Sigma^{-1} e}.$$

3. Risikobasierte Portfoliooptimierung

Da aus der positiven Definitheit von Σ folgt, dass auch Σ^{-1} positiv definit ist, gilt stets $\gamma > 0$. Wir erhalten damit als eindeutiges Optimum

$$w^* = \frac{\Sigma^{-1}e}{e^\top \Sigma^{-1}e}. \quad (3.2)$$

Die einzelnen Assetgewichte können in dieser Lösung durchaus auch negative Werte annehmen.

Als nächstes möchten wir noch den Spezialfall konstanter Asset-Korrelationen (d.h. $\varrho_{i,j} = \varrho, \forall i \neq j$) für dieses Problem genauer behandeln:

Sei $D := \text{diag}(\sigma_1, \dots, \sigma_n)$ die Diagonalmatrix der Standardabweichungen der einzelnen Asset>Returns und $P_n(\varrho) = P = (\varrho_{i,j})_{1 \leq i,j \leq n}$ die Korrelationsmatrix mit $\varrho_{i,j} = \varrho$ für $i \neq j$ sowie $\varrho_{i,i} = 1$. Dann gilt für die Kovarianzmatrix $\Sigma = DPD$ und für deren Inverse $\Sigma^{-1} = D^{-1}P^{-1}D^{-1}$ mit

$$P^{-1} = \frac{(1 + (n-1)\varrho)I_n - \varrho \cdot ee^\top}{1 + (n-2)\varrho - (n-1)\varrho^2}.$$

Die genaue Berechnung dieser Matrix ist in Appendix A.3 zu finden. Mit (3.2) erhalten wir insgesamt

$$w_i^* = \frac{(1 + (n-1)\varrho)\sigma_i^{-2} - \varrho \sum_{j=1}^n (\sigma_i\sigma_j)^{-1}}{\sum_{k=1}^n \left((1 + (n-1)\varrho)\sigma_k^{-2} - \varrho \sum_{j=1}^n (\sigma_k\sigma_j)^{-1} \right)} \quad (3.3)$$

als Optimallösung von (\overline{MV}) , wobei nicht ausgeschlossen ist, dass auch negative Assetgewichte auftreten.

Die Korrelation ϱ muss im Intervall $[-\frac{1}{n-1}, 1]$ liegen, damit $P_n(\varrho)$ eine zulässige (also positiv semidefinite) Korrelationsmatrix ist (siehe ebenfalls Appendix A.3). An den Rändern dieses Bereiches ist $P_n(\varrho)$ jedoch nicht mehr positiv definit und die obigen Berechnungen (mit der inversen Korrelations- bzw. Kovarianzmatrix) sind nicht mehr durchführbar. Wir untersuchen den Fall $\varrho = -\frac{1}{n-1}$ etwas genauer, da hierfür ein interessanter Zusammenhang mit dem „Risk-Parity“-Ansatz, den wir später kennenlernen werden, besteht. Anstatt der positiven Definitheit von Σ bzw. P nehmen wir nun nur mehr an, dass die Standardabweichungen σ_i alle echt größer als 0 sind. Dies schließt das Vorhandensein eines risikolosen Assets aus und impliziert die Invertierbarkeit von D .²

Wir betrachten nochmals die KKT-Bedingung $\Sigma w = \gamma e$. Mit $\tilde{w} := Dw$ (und $\Sigma = DPD$) entspricht das der Forderung

$$P\tilde{w} = \gamma \cdot D^{-1}e.$$

²Für Modelle mit risikolosem Asset hat das Problem (\overline{MV}) eine triviale Optimallösung mit Gewicht 1 für ebendieses, Gewicht 0 für die übrigen risikobehafteten Assets und optimalem Zielfunktionswert 0.

3. Risikobasierte Portfoliooptimierung

Wählen wir $\gamma = 0$ und $\tilde{w} = (t, \dots, t)^\top$ als konstanten Vektor, so ist dies offenbar erfüllt, da

$$(P\tilde{w})_i = t \cdot \left(1 - (n-1) \frac{1}{n-1}\right) = 0.$$

Mit $w = D^{-1}\tilde{w} = (t\sigma_1^{-1}, \dots, t\sigma_n^{-1})^\top$ und der Normierungsbedingung $\sum_{k=1}^n w_k = 1$ folgt $t = \frac{1}{\sum_{k=1}^n \sigma_k^{-1}}$ und die Optimallösung

$$w_i^* = \frac{\sigma_i^{-1}}{\sum_{k=1}^n \sigma_k^{-1}}.$$

Dies entspricht allerdings wieder genau dem Ausdruck in (3.3) mit $\varrho = -\frac{1}{n-1}$. Man beachte, dass hier das optimale Portfolio für (\overline{MV}) über positive Gewichte für alle Assets verfügt und somit auch ein Optimum von (MV) ist.

3.2. Der „Risk-Budgeting“- und „Risk-Parity“-Ansatz

Zwar führt die Herangehensweise im „Minimum-Variance“-Modell zu gleichen Risikobeiträgen, allerdings nur auf marginaler Ebene. Im „Risk-Parity“-Ansatz spielt nun der totale Risikobeitrag eine entscheidende Rolle. Dieser ist folgendermaßen definiert:

Definition. ((Totaler) Risikobeitrag, vgl. Bai, Scheinberg und Tütüncü (2013))

Betrachte eine Menge von n Assets mit Kovarianzmatrix Σ und eine Menge Ω zulässiger Portfolios. Für ein $w \in \Omega$ (mit Portfoliorisiko $w^\top \Sigma w = 2 \cdot \zeta^2(w)$) ist der **(totale) Risikobeitrag** des i -ten Assets definiert als

$$\zeta_i^2(w) := w_i \cdot \partial_{w_i} \zeta^2(w) = w_i (\Sigma w)_i = w_i^2 \sigma_i^2 + \sum_{\substack{j=1 \\ j \neq i}}^n w_i w_j \sigma_{i,j}$$

und entspricht somit dem Produkt des marginalen Risikobeitrags mit dem entsprechenden Assetgewicht.

Den Zusammenhang zwischen dem Gesamtrisiko eines Portfolios und dem Risikobeitrag der einzelnen Assets gibt nun folgende Zerlegung, welche auch die Bezeichnung der ζ_i^2 als *totale* Risikobeiträge rechtfertigt, an:

$$w^\top \Sigma w = 2 \cdot \zeta^2(w) = \sum_{i=1}^n \zeta_i^2(w) \quad (3.4)$$

Das Gesamtrisiko eines Portfolios ist also die Summe der totalen Risikobeiträge der einzelnen Assets. Diese Zerlegung folgt unmittelbar aus der Definition des marginalen und totalen

3. Risikobasierte Portfoliooptimierung

Risikobeitrags bzw. lässt sich aus dem Satz von Euler für positiv homogene Funktionen ableiten, da $\zeta^2(w)$ eine positiv homogene Funktion vom Grad zwei ist (siehe Appendix A.2). Wir werden in weiterer Folge für ein Portfolio w den Wert $\frac{\zeta_i^2(w)}{w^\top \Sigma w}$ als den **(totalen) Risikoanteil** des i -ten Assets am Portfoliorisiko bezeichnen.

Um zu erreichen, dass jedes Assets den gleichen Beitrag zur Portfolio-Varianz leistet, müssen die totalen Risikobeiträge gleichgesetzt werden:

Definition. („Risk-Parity“-Problem / „Risk-Parity“-Lösung)

Sei Ω eine Menge zulässiger Portfolios. Dann nennen wir das Problem, ein $w^* \in \Omega$ zu finden, sodass die Bedingung

$$(RP) \quad \zeta_i^2(w^*) = \zeta_j^2(w^*), \forall i, j \in \{1, \dots, n\}$$

erfüllt ist, das „**Risk-Parity**“-Problem.

Jede zulässige Lösung von (RP) heißt „**Risk-Parity**“-Lösung oder „**Risk-Parity**“-Portfolio bezüglich Ω .

Die Menge aller zulässigen „Risk-Parity“-Lösungen bezüglich Ω bezeichnen wir mit $\mathcal{RP}(\Omega)$.

Dabei ist zu beachten, dass die Menge der „Risk-Parity“-Lösungen auch leer sein kann, da es nicht zwangsläufig ein zulässiges Portfolio geben muss, welches im „Risk-Parity“-Zustand (RP) ist. In diesem Fall nennen wir das Problem unzulässig.

Aus (3.4) folgt direkt, dass in einer „Risk-Parity“-Lösung w^* der totale Risikobeitrag jedes Assets $\frac{w^{*\top} \Sigma w^*}{n}$ (und somit jeder Risikoanteil $1/n$) beträgt. Da die Portfolio-Varianz nichtnegativ ist, sind somit auch alle totalen Risikobeiträge stets nichtnegativ.

Hat ein Investor spezielle Präferenzen für einzelne Assets, so ist es ihm jedoch nicht möglich, diese in die Formulierung des obigen Problems einfließen zu lassen. Daher definieren wir nun noch das „Risk-Budgeting“-Problem, in dem für jedes Asset beliebig vorgegeben werden kann, welchen Risikoanteil es aufweisen soll. Wir orientieren uns dabei an der entsprechenden Definition in Roncalli (2013):

Definition. („Risk-Budgeting“-Problem / „Risk-Budgeting“-Lösung)

Sei Ω eine Menge zulässiger Portfolios. Dann nennen wir das Problem, für einen Vektor $b = (b_1, \dots, b_n) \in \mathbb{R}^n$ mit $b_i > 0, \forall i = 1, \dots, n$ und $\sum_{i=1}^n b_i = 1$ ein Portfolio $w^* \in \Omega$ zu finden, sodass die Bedingung

$$(RB_b) \quad \frac{\zeta_i^2(w^*)}{w^{*\top} \Sigma w^*} = b_i, \forall i \in \{1, \dots, n\}$$

erfüllt ist, das „**Risk-Budgeting**“-Problem.

Jede zulässige Lösung von (RB_b) heißt „**Risk-Budgeting**“-Lösung oder „**Risk-Budgeting**“-Portfolio bezüglich b und Ω .

Die Menge aller zulässigen „Risk-Budgeting“-Lösungen bezüglich b und Ω bezeichnen wir mit $\mathcal{RB}_b(\Omega)$.

3. Risikobasierte Portfoliooptimierung

Damit entspricht der „Risk-Parity“-Ansatz also dem Spezialfall dieses Problems mit Risikoanteilen $b_i = \hat{b}_i := 1/n, \forall i = 1, \dots, n$ und es gilt $\mathcal{RP}(\Omega) = \mathcal{RB}_{\hat{b}}(\Omega)$.

Wir haben uns bei unserer Definition auf strikt positive Risikoanteile beschränkt. Fordern wir hingegen nur nichtnegative Risikoanteile, so besteht *eine* Möglichkeit für die Berechnung eines „Risk-Budgeting“-Portfolios darin, das Problem nur für die Menge der Assets mit positiven Risikoanteilen zu lösen und $w_i = 0$ für jedes Asset mit $b_i = 0$ zu setzen. Wie wir nun sehen werden, ist dies jedoch sogar im „Long-Only“-Fall im Allgemeinen nicht die einzige Lösung (vgl. Roncalli, 2013):

Sei $\mathbb{R}_{I_n,1}^n := \{w \in \mathbb{R}^n \mid I_n w \geq 0, e^\top w = 1\}$ die Zulässigkeitsmenge und der Risikoanteil $b_k = 0$. Wir erhalten folgende Bedingung an den Risikobeitrag des k -ten Assets:

$$\varsigma_k^2(w) = w_k^2 \sigma_k^2 + \sum_{\substack{j=1 \\ j \neq k}}^n w_k w_j \sigma_{k,j} = w_k \left(w_k \sigma_k^2 + \sigma_k \sum_{\substack{j=1 \\ j \neq k}}^n w_j \varrho_{k,j} \sigma_j \right) = 0$$

Diese Gleichung besitzt zwei Lösungen:

- $w_k^{(1)} = 0$
- $w_k^{(2)} = -\frac{1}{\sigma_k} \sum_{\substack{j=1 \\ j \neq k}}^n w_j \varrho_{k,j} \sigma_j$

Falls $\varrho_{k,j} \geq 0, \forall j = 1, \dots, n$, so folgt $w_k^{(2)} \leq 0$, womit $w_k^{(1)} = 0$ die einzig mögliche zulässige Lösung ist. Treten hingegen auch negative Korrelationen auf, so ist die Existenz einer Lösung mit $w_k > 0$ nicht ausgeschlossen.

Roncalli (2013) hat diesen Fall genauer untersucht und unter anderem festgestellt, dass es eine Lösung mit $w_k > 0$ genau dann gibt, wenn der marginale Risikobeitrag des k -ten Assets im „Risk-Budgeting“-Portfolio mit $w_k = 0$ *negativ* ist.

Wir beschränken uns im Folgenden auf den Fall, dass die Risikoanteile aller Assets strikt positiv sind.

3.2.1. Ein Lösungsversuch des „Risk-Parity“-Problems

Sei $\mathbb{R}_1^n := \{w \in \mathbb{R}^n \mid e^\top w = 1\}$ die Menge der zulässigen Portfolios für das „Risk-Parity“-Problem. Die einzige Restriktion ist also wieder (wie bei der Lösung des „Minimum-Variance“-Modells in Abschnitt 3.1.1) die Normierung der Gewichte, sodass deren Summe 1 ergibt. Wir möchten nun zwei Spezialfälle dieses Problems untersuchen (vgl. wieder Maillard, Roncalli und Teiletche (2010) sowie Roncalli (2013)), wobei wir noch voraussetzen, dass die

3. Risikobasierte Portfoliooptimierung

Standardabweichungen aller Assets positiv sind:³

Konstante Korrelationen

Wir nehmen an, dass die Asset-Korrelationen konstant (d.h. $\rho_{i,j} = \rho, \forall i \neq j$ mit $\rho \in [-\frac{1}{n-1}, 1]$, vgl. Appendix A.3) sind.

Dann gilt für den totalen Risikobeitrag des i -ten Assets für ein beliebiges Portfolio $w \in \mathbb{R}_1^n$

$$\varsigma_i^2(w) = w_i^2 \sigma_i^2 + \sum_{\substack{k=1 \\ k \neq i}}^n w_i w_k \rho \sigma_i \sigma_k = w_i \sigma_i \left((1 - \rho) w_i \sigma_i + \rho \sum_{k=1}^n w_k \sigma_k \right).$$

Somit folgt:

$$\begin{aligned} \varsigma_i^2(w) &= \varsigma_j^2(w) \\ &\Leftrightarrow \\ w_i \sigma_i \left((1 - \rho) w_i \sigma_i + \rho \sum_{k=1}^n w_k \sigma_k \right) &= w_j \sigma_j \left((1 - \rho) w_j \sigma_j + \rho \sum_{k=1}^n w_k \sigma_k \right) \end{aligned}$$

Diese Gleichung ist offensichtlich erfüllt, wenn $w_i \sigma_i = w_j \sigma_j \Leftrightarrow w_j = w_i \frac{\sigma_i}{\sigma_j}$ für alle $j = 1, \dots, n$ ist. Wir verwenden nun die Normierungsbedingung $e^\top w = 1$ und erhalten unmittelbar eine (nicht notwendigerweise eindeutige) Optimallösung w^* :

$$1 = w_i + \sum_{\substack{j=1 \\ j \neq i}}^n w_j = w_i + \sum_{\substack{j=1 \\ j \neq i}}^n w_i \frac{\sigma_i}{\sigma_j} = w_i \sigma_i \sum_{j=1}^n \sigma_j^{-1} \quad \Rightarrow \quad w_i^* = \frac{\sigma_i^{-1}}{\sum_{j=1}^n \sigma_j^{-1}}$$

Das Gewicht eines Assets entspricht also seiner normierten reziproken Standardabweichung. Risikoreiche Assets bekommen damit entsprechend niedrigere Gewichte als risikoarme.

Es ist hier zu beachten, dass dieses Portfolio unabhängig von der Wahl von ρ immer optimal ist und ausschließlich über positive Assetgewichte verfügt. Für den Spezialfall $\rho = -\frac{1}{n-1}$ stimmt die obige „Risk-Parity“-Lösung mit der optimalen Lösung des „Minimum-Variance“-Problems (\overline{MV}) (siehe Kapitel 3.1.1) überein.

Konstante Varianzen

Nun versuchen wir ohne Voraussetzungen an die Korrelationen, jedoch mit konstanten Varianzen (d.h. $\sigma_i^2 = \sigma^2 > 0, \forall i = 1, \dots, n$), eine „Risk-Parity“-Lösung zu berechnen.

³Man beachte, dass ein Asset mit einer Standardabweichung von 0 (dies entspricht einer risikolosen Anlagemöglichkeit) wie beim „Minimum-Variance“-Problem zu einer Optimallösung führt, in der die gesamte Summe in ebendieses Asset investiert wird. Wir fordern nämlich gleiche Risikobeiträge für alle Assets und der Risikobeitrag des risikolosen Assets ist (unabhängig von der Gewichtung) immer gleich 0, während alle risikobehafteten Assets ebenfalls einen Risikobeitrag von 0 aufweisen, wenn sie mit 0 gewichtet werden.

3. Risikobasierte Portfoliooptimierung

Mit dieser Annahme erhalten wir für ein beliebiges Portfolio $w \in \mathbb{R}_1^n$ als totale Risikobeiträge

$$\zeta_i^2(w) = w_i^2 \sigma_i^2 + \sum_{\substack{k=1 \\ k \neq i}}^n w_i w_k \varrho_{i,k} \sigma_i \sigma_k = w_i \sigma^2 \sum_{k=1}^n w_k \varrho_{i,k}. \quad (3.5)$$

Dies liefert uns folgende Aussage:

$$\begin{aligned} \zeta_i^2(w) &= \zeta_j^2(w) \\ \Leftrightarrow \\ w_i \sum_{k=1}^n w_k \varrho_{i,k} &= w_j \sum_{k=1}^n w_k \varrho_{j,k} \end{aligned}$$

Daraus folgt $w_j = w_i \frac{\sum_{k=1}^n w_k \varrho_{i,k}}{\sum_{k=1}^n w_k \varrho_{j,k}}$. Setzen wir schlussendlich in die Bedingung $e^\top w = 1$ ein, so erhalten wir nachstehendes Resultat:

$$1 = w_i + \sum_{\substack{j=1 \\ j \neq i}}^n w_j = w_i + \sum_{\substack{j=1 \\ j \neq i}}^n w_i \frac{\sum_{k=1}^n w_k \varrho_{i,k}}{\sum_{k=1}^n w_k \varrho_{j,k}} = w_i \sum_{j=1}^n \frac{\sum_{k=1}^n w_k \varrho_{i,k}}{\sum_{k=1}^n w_k \varrho_{j,k}} = w_i \frac{\sum_{j=1}^n (\sum_{k=1}^n w_k \varrho_{j,k})^{-1}}{(\sum_{k=1}^n w_k \varrho_{i,k})^{-1}}$$

Es gilt also

$$w_i = \frac{(\sum_{k=1}^n w_k \varrho_{i,k})^{-1}}{\sum_{j=1}^n (\sum_{k=1}^n w_k \varrho_{j,k})^{-1}}. \quad (3.6)$$

Unter Verwendung dieser Darstellung ist es für $n \geq 3$ nicht möglich, das Problem in geschlossener Form zu lösen.⁴ Die Annahme konstanter Varianzen liefert also im Gegensatz zu konstanten Korrelationen nicht unmittelbar eine explizite Lösung.

Alternativ zu (3.6) können wir durch Gleichsetzen von (3.5) und $\zeta_i^2(w) = \frac{w^\top \Sigma w}{n}$ auch direkt

$$w_i = w^\top \Sigma w \cdot \frac{(\sum_{k=1}^n w_k \varrho_{i,k})^{-1}}{n \sigma^2}$$

als Bedingung an eine „Risk-Parity“-Lösung angeben. Führen wir obige Berechnungen ohne konstante Varianzen aus, so entspricht dies der Vorschrift

$$w_i = w^\top \Sigma w \cdot \frac{(\sum_{k=1}^n w_k \sigma_{i,k})^{-1}}{n}.$$

Daraus können wir eine interessante und wünschenswerte Eigenschaft ablesen: Besitzt ein Asset eine hohe Varianz und/oder ist es stark mit anderen Assets korreliert, so führt dies zu einem geringeren Gewicht in einer „Risk-Parity“-Lösung.

⁴Für $n = 2$ gibt es nur einen Korrelationskoeffizienten $\varrho_{1,2} = \varrho_{2,1}$, womit der Spezialfall mit konstanten Korrelationen anwendbar ist.

3.3. „Risk-Budgeting“-Portfolios als Lösungen nichtlinearer Optimierungsprobleme

In Abschnitt 3.2.1 ist uns die Berechnung eines „Risk-Parity“-Portfolios in expliziter Form nur unter der Annahme konstanter Korrelationen gelungen. Eine analoge Vorgangsweise für das „Risk-Budgeting“-Problem liefert explizite Lösungen sogar nur für die Spezialfälle $\varrho = 0$, $\varrho = 1$ und $\varrho = -1/n-1$ (siehe Roncalli, 2013). Klarerweise ist diese Voraussetzung in der Praxis im Allgemeinen nicht erfüllt. Möchte man beispielsweise in die Aktien einer gewissen Anzahl von Unternehmen aus verschiedenen Branchen investieren, so korrelieren zwei Aktien in der Regel deutlich stärker, wenn die entsprechenden Unternehmen in derselben Branche tätig sind, als wenn sie aus unterschiedlichen Branchen stammen. Daher möchten wir nun nichtlineare Optimierungsprobleme in der Form (NLP) betrachten, mithilfe derer es möglich ist, „Risk-Budgeting“-Lösungen für beliebige Kovarianzmatrizen aufzuspüren.

Wir treffen folgende Annahme:

Die Kovarianzmatrix Σ sei positiv definit.

Dies lässt sich folgendermaßen rechtfertigen:

- Seien alle betrachteten Assets risikobehaftet. Es gelte also für die Standardabweichungen der Asset>Returns $\sigma_i > 0$, $\forall i = 1, \dots, n$. Dann folgt für eine positiv semidefinite, jedoch nicht positiv definite Kovarianzmatrix Σ die Existenz eines Vektors $w \in \mathbb{R}^n \setminus \{0\}$ mit

$$w^\top \Sigma w = \text{Var}(R(w)) = \text{Var}(w_1 R_1 + \dots + w_n R_n) = 0.$$

Dies bedeutet jedoch, dass $R(w)$ fast sicher konstant und damit die zufälligen Asset>Returns fast sicher linear abhängig sind. Diesen Fall möchten wir ausschließen.

- Seien nun die ersten $n - 1$ Assets risikobehaftet mit fast sicher linear unabhängigen Returns und das n -te Asset risikolos. Dann lässt sich die Kovarianzmatrix $\Sigma \in \mathbb{R}^{n \times n}$ wie folgt darstellen:

$$\Sigma = \begin{pmatrix} \widehat{\Sigma} & 0 \\ 0^\top & 0 \end{pmatrix}$$

wobei $\widehat{\Sigma} \in \mathbb{R}^{(n-1) \times (n-1)}$ die positiv definite Kovarianzmatrix der risikobehafteten Assets ist. Der totale Risikobeitrag des risikolosen Assets ist bei beliebiger Gewichtung immer gleich 0. Falls $w \in \mathbb{R}^n$ eine „Risk-Budgeting“-Lösung bezüglich $b > 0$ ist, so ist also der Risikobeitrag $b_n \cdot w^\top \Sigma w = 0$, womit $w^\top \Sigma w = 0$ folgt. Andererseits ist $w^\top \Sigma w = \widehat{w}^\top \widehat{\Sigma} \widehat{w}$ mit $\widehat{w} = (w_1, \dots, w_{n-1})^\top \in \mathbb{R}^{n-1}$. Aufgrund der positiven Definitheit von $\widehat{\Sigma}$ resultiert $\widehat{w} = (0, \dots, 0)^\top$. Wird die Menge der zulässigen Portfolios beispielsweise noch durch die Normierungsrestriktion $w^\top e = 1$ eingeschränkt, so ist $w = (0, \dots, 0, 1)^\top$ das *eindeutige* „Risk-Budgeting“-Portfolio bezüglich b . Daher möchten wir auch keine risikolosen Assets in unseren weiteren Analysen berücksichtigen.

3. Risikobasierte Portfoliooptimierung

3.3.1. Ein konvexes Optimierungsproblem mit logarithmischer Schranke

Wir untersuchen zu Beginn ein Optimierungsproblem mit einer logarithmischen Schranke in der Zielfunktion. Diese Herangehensweise wird unter anderem in Bai, Scheinberg und Tütüncü (2013) sowie Spinu (2013) gewählt. Maillard, Roncalli und Teiletche (2010), Bruder und Roncalli (2012), Roncalli (2013) sowie Kaya und Lee (2012) arbeiten mit ähnlichen Optimierungsproblemen. In den ersten drei Papers wird beispielsweise die logarithmische Schranke in Form einer Restriktion (anstatt als Teil der Zielfunktion) verwendet.

Wir werden die Notationen

- $\mathbb{R}_+^n := \{w \in \mathbb{R}^n \mid w > 0\}$,
- $\mathbb{R}_{+,1}^n := \{w \in \mathbb{R}^n \mid w > 0, e^\top w = 1\}$,
- $\mathbb{R}_A^n := \{w \in \mathbb{R}^n \mid Aw \geq 0\}$,
- $\mathbb{R}_{A,1}^n := \{w \in \mathbb{R}^n \mid Aw \geq 0, e^\top w = 1\}$,
- $\mathbb{R}_1^n := \{w \in \mathbb{R}^n \mid e^\top w = 1\}$,

verwenden, wobei A eine reelle $(n \times n)$ -Matrix ist.

Des Weiteren benötigen wir folgende Eigenschaft einer „Risk-Budgeting“-Lösung:

Lemma 3.3.1. *Sei $\Sigma \succ 0$, $b \in \mathbb{R}_{+,1}^n$, \mathbb{R}^n die Menge der zulässigen Portfolios und $w^* \in \mathcal{RB}_b(\mathbb{R}^n)$. Dann $\nexists i \neq j \in \{1, \dots, n\}$, sodass $w_i^* = 0$ und $w_j^* \neq 0$.*

Beweis. Sei $w^* \in \mathcal{RB}_b(\mathbb{R}^n)$ mit $w_i^* = 0$. Daraus folgt für den Risikobeitrag des i -ten Assets

$$\zeta_i^2(w^*) = w_i^* \cdot \partial_{w_i} \zeta^2(w) \Big|_{w=w^*} = b_i \cdot w^{*\top} \Sigma w^* = 0,$$

womit wegen $b_i > 0$ das Portfoliorisiko $w^{*\top} \Sigma w^*$ ebenfalls gleich 0 ist. Da Σ laut Annahme positiv definit ist, folgt $w^* = 0$. \square

Wir betrachten nun schrittweise verschiedene Zulässigkeitsmengen:

Im ersten Schritt beschränken wir uns auf „Long-Only“-Portfolios ohne Gewichtsnormierung. Unsere Zulässigkeitsmenge hat also die Form $\mathbb{R}_{I_n}^n = \{w \in \mathbb{R}^n \mid w \geq 0\}$. Für eine „Risk-Budgeting“-Lösung $w^* \in \mathcal{RB}_b(\mathbb{R}_{I_n}^n)$ muss dann wegen Lemma 3.3.1 eine der folgenden Aussagen gelten:

- $w^* = 0$. Alle Risikobeiträge und somit auch das Portfoliorisiko sind gleich 0.
- $w^* \in \mathbb{R}_+^n$. Alle Risikobeiträge sind positiv.

3. Risikobasierte Portfoliooptimierung

Wir interessieren uns für den zweiten Fall und betrachten folgendes Optimierungsproblem mit $\mathbb{R}_{I_n}^n$ als Zulässigkeitsmenge:

$$\begin{aligned}
 (\text{LS}^c) \quad & \min_{w \in \mathbb{R}^n} \quad \frac{1}{2} w^\top \Sigma w - c \sum_{i=1}^n b_i \ln w_i \\
 & \text{s.t.} \quad w_i \geq 0, \quad i = 1, \dots, n
 \end{aligned}$$

Der Parameter c ist dabei eine beliebige positive Konstante. Die Zielfunktion, welche wir als Abbildung von $\mathbb{R}_{I_n}^n$ nach $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ betrachten (mit $\ln 0 = -\infty$), ist strikt konvex, da einerseits Σ laut Annahme positiv definit (und damit $w^\top \Sigma w$ strikt konvex) und andererseits der Logarithmus eine strikt konkave Funktion ist. Da alle Restriktionen linear sind, erhalten wir insgesamt ein konvexes Optimierungsproblem (KP). Gelingt es uns, eine Lösung von (LS^c) (für ein festes $c > 0$) zu finden, so wissen wir aufgrund der strikten Konvexität der Zielfunktion, dass dieses Minimum eindeutig ist. Wie wir nun sehen werden, ist auch die Existenz einer Optimallösung für obiges Problem gegeben:

Lemma 3.3.2. *Sei $\Sigma \succ 0$ und $b \in \mathbb{R}_{+,1}^n$. Dann existiert für jedes beliebige $c > 0$ ein eindeutiges globales Minimum $w^{(c)} \in \mathbb{R}_{I_n}^n$ von (LS^c) , wobei sogar $w^{(c)} \in \mathbb{R}_+^n$ gilt.*

Beweis. (Siehe Spinu (2013). Vgl. Kanzow (2005) für die verwendeten Aussagen aus der linearen Algebra.)

Sei

$$\begin{aligned}
 f_{\text{LS}}^c(w) : \mathbb{R}_{I_n}^n &\rightarrow \overline{\mathbb{R}} \\
 w &\mapsto \frac{1}{2} w^\top \Sigma w - c \sum_{i=1}^n b_i \ln w_i
 \end{aligned}$$

die strikt konvexe Zielfunktion von (LS^c) mit beliebigem, festem $c > 0$. Wir zeigen nun

$$\lim_{w \rightarrow \partial \mathbb{R}_{I_n}^n} f_{\text{LS}}^c(w) = +\infty, \tag{3.7}$$

woraus unmittelbar die Aussage des Satzes folgt:

Die positiv definite Kovarianzmatrix Σ ist reell und symmetrisch und damit normal (d.h. $\Sigma^\top \Sigma = \Sigma \Sigma^\top$). Wegen dieser Eigenschaft ist Σ unitär diagonalisierbar mit $U^\top \Sigma U = D$. U ist dabei eine reguläre, orthogonale Matrix, womit $U^\top = U^{-1}$ gilt, und $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ eine Diagonalmatrix mit den (positiven) Eigenwerten $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ von Σ auf der Diagonalen. Dann gilt also für einen Vektor $w \in \mathbb{R}_{I_n}^n$ und $v := U^\top w$ (und damit $w = Uv$),

3. Risikobasierte Portfoliooptimierung

dass

$$\begin{aligned} w^\top \Sigma w &= v^\top U^\top \Sigma U v = v^\top U^\top U D U^\top U v = v^\top U^{-1} U D U^{-1} U v = v^\top D v = \sum_{i=1}^n \lambda_i v_i^2 \\ &\geq \lambda_1 \sum_{i=1}^n v_i^2 = \lambda_1 v^\top v = \lambda_1 w^\top U U^\top w = \lambda_1 w^\top U U^{-1} w = \lambda_1 w^\top w \\ &= \lambda_1 \sum_{i=1}^n w_i^2. \end{aligned}$$

Daher gilt

$$f_{\text{LS}}^c(w) \geq \sum_{i=1}^n \left(\frac{1}{2} \lambda_1 w_i^2 - c b_i \ln w_i \right) = \sum_{i=1}^n \varphi_i(w_i),$$

mit $\varphi_i(x) : \mathbb{R}_0^+ \rightarrow \overline{\mathbb{R}}$; $x \mapsto \frac{1}{2} \lambda_1 x^2 - c b_i \ln x$.⁵ Die strikt konvexe Funktion φ_i hat im Punkt $x_i^* = \sqrt{\frac{c b_i}{\lambda_1}} > 0$ ein globales Minimum, weshalb wir für jedes $i \in \{1, \dots, n\}$

$$f_{\text{LS}}^c(w) \geq \varphi_i(w_i) + \sum_{\substack{j=1 \\ j \neq i}}^n \varphi_j(x_j^*) \quad (3.8)$$

folgern können. Des Weiteren erhalten wir

$$\lim_{x \rightarrow +\infty} \varphi_i(x) = +\infty \quad \text{und} \quad \lim_{x \rightarrow 0^+} \varphi_i(x) = +\infty, \quad (3.9)$$

für alle $i = 1, \dots, n$, da einerseits x^2 für $x \rightarrow +\infty$ schneller wächst als $\ln x$ und andererseits $\lim_{x \rightarrow 0^+} \ln x = -\infty$ ist.

Sei nun $M > 0$ beliebig. Aus (3.8) und (3.9) folgt, dass es zwei Zahlen c_i^M und $d_i^M \in (0, +\infty)$ gibt, sodass $f_{\text{LS}}^c(w) \geq M$ für alle $w \in \mathbb{R}_{I_n}^n$ mit $w_i \notin [c_i^M, d_i^M]$. Daher gilt

$$w \notin [c_1^M, d_1^M] \times [c_2^M, d_2^M] \times \dots \times [c_n^M, d_n^M] \quad \Rightarrow \quad f_{\text{LS}}^c(w) \geq M.$$

Das beweist (3.7). □

Folgender Satz gibt nun den Zusammenhang zwischen dem Programm (LS^c) und den „Risk-Budgeting“-Portfolios mit positiven Assetgewichten an und zeigt uns die Bedeutung des Parameters c :

Satz 3.3.3. *Sei die Kovarianzmatrix $\Sigma \succ 0$ und $b \in \mathbb{R}_{+,1}^n$. Dann gelten folgende Aussagen:*

- a) *Sei $c > 0$ beliebig. Ein Portfolio $w^{(c)} \in \mathbb{R}_{I_n}^n$ ist genau dann eine Optimallösung von (LS^c), wenn $w^{(c)}$ eine „Risk-Budgeting“-Lösung bezüglich b mit totalen Risikobeiträgen $\zeta_i^2(w^{(c)}) = b_i c$ für alle $i = 1, \dots, n$ ist.*

⁵ $\mathbb{R}_0^+ = \{x \in \mathbb{R} \mid x \geq 0\} = \mathbb{R}_{I_1}^1$ in der von uns eingeführten Notation.

3. Risikobasierte Portfoliooptimierung

b) Es existiert für jedes beliebige $c > 0$ genau ein „Risk-Budgeting“-Portfolio $w^{(c)} \in \mathbb{R}_{I_n}^n$ bezüglich b mit Risikobeiträgen $\zeta_i^2(w^{(c)}) = b_i c, \forall i = 1, \dots, n$, wobei $w^{(c)} \in \mathbb{R}_+^n$ gilt.

Beweis.

a) „ \Rightarrow “: Ist $w^{(c)}$ ein Minimum von (LS^c) mit $c > 0$, so gibt es einen KKT-Punkt $(w^{(c)}, \lambda) \in \mathbb{R}^n \times \mathbb{R}^n$, der die KKT-Bedingungen

$$\begin{aligned} [LS^c-i] \quad & w^{(c)} \geq 0, \\ [LS^c-ii] \quad & (\Sigma w^{(c)})_i - b_i c / w_i^{(c)} - \lambda_i = 0, \forall i = 1, \dots, n, \\ [LS^c-iii] \quad & \lambda \geq 0, \lambda^\top w^{(c)} = 0, \end{aligned}$$

erfüllt (siehe Korollar 1.4.8). Dabei sind die einzelnen Gewichte $w_i^{(c)} > 0$ für alle $i = 1, \dots, n$ (vgl. Lemma 3.3.2). Somit gilt aufgrund der Bedingung $[LS^c-iii]$ auf jeden Fall $\lambda = 0$. Setzen wir in $[LS^c-ii]$ ein, so folgt

$$\begin{aligned} (\Sigma w^{(c)})_i &= \frac{b_i c}{w_i^{(c)}}, \forall i \in \{1, \dots, n\} \\ &\Leftrightarrow \\ \zeta_i^2(w^{(c)}) &= b_i c, \forall i, j \in \{1, \dots, n\} \end{aligned}$$

womit $w^{(c)} \in \mathcal{RB}_b(\mathbb{R}_{I_n}^n)$ (mit totalen Risikobeiträgen $b_i c$ für die einzelnen Assets) ist.

„ \Leftarrow “: Sei $w^{(c)} \in \mathbb{R}_{I_n}^n$ eine „Risk-Budgeting“-Lösung mit *positiven* Risikobeiträgen $\zeta_i^2(w^{(c)}) = b_i c, \forall i = 1, \dots, n$. Dann sind alle Komponenten von $w^{(c)}$ wegen $\Sigma \succ 0$ und Lemma 3.3.1 positiv ($w_i^{(c)} > 0, \forall i = 1, \dots, n$).

Der Punkt $(w^{(c)}, 0) \in \mathbb{R}^n \times \mathbb{R}^n$ ist nun offensichtlich ein KKT-Punkt von (LS^c) und damit ein Minimum dieses konvexen Optimierungsproblems (siehe Korollar 1.4.8).

b) Folgt direkt aus a) und Lemma 3.3.2. □

Somit wissen wir nun, dass (unter den Voraussetzungen des obigen Satzes) die Menge der „Risk-Budgeting“-Lösungen nicht leer ist. Nur wenn zu starke Restriktionen an die Assetgewichte gestellt werden, kann der Fall eintreten, dass es kein „Risk-Budgeting“-Portfolio gibt.

Eine oft geforderte Bedingung ist beispielsweise die Normierung der Gewichte, sodass ihre Summe gleich 1 ergibt, was jedoch unproblematisch ist: Das folgende Korollar besagt nämlich, dass diese Restriktion einerseits keinen Einfluss auf obige Existenzaussage hat und andererseits sogar zur Eindeutigkeit eines „Risk-Budgeting“-Portfolios führt.

Korollar 3.3.4. Sei $\Sigma \succ 0$ und $b \in \mathbb{R}_{+,1}^n$. Dann existiert genau ein „Risk-Budgeting“-Portfolio $w^* \in \mathbb{R}_{I_n,1}^n$ bezüglich b .

3. Risikobasierte Portfoliooptimierung

Beweis. (Vgl. Bai, Scheinberg und Tütüncü (2013) für den „Risk-Parity“-Spezialfall.) Sei $c = c_1 > 0$ gegeben. Dann existiert nach Lemma 3.3.2 bzw. Satz 3.3.3 eine eindeutige Lösung $w^{(c_1)}$ von (LS^{c_1}) mit $\zeta_i^2(w^{(c_1)}) = b_i c_1, \forall i = 1, \dots, n$. Wir betrachten nun für eine positive Konstante a das Optimierungsproblem (LS^{ac_1}) , also

$$\begin{aligned} \min_{w \in \mathbb{R}^n} \quad & \frac{1}{2} w^\top \Sigma w - ac_1 \sum_{i=1}^n b_i \ln w_i \\ \text{s.t.} \quad & w_i \geq 0, \quad i = 1, \dots, n \end{aligned}$$

Das (eindeutige) Optimum dieses Problems ist $\sqrt{aw}^{(c_1)}$. Für die Risikobeiträge des i -ten Assets gilt in diesem Portfolio nämlich

$$\zeta_i^2(\sqrt{aw}^{(c_1)}) = \sqrt{aw}^{(c_1)} (\Sigma \sqrt{aw}^{(c_1)})_i = a \cdot w_i^{(c_1)} (\Sigma w^{(c_1)})_i = b_i ac_1.$$

Da alle Komponenten von $w^{(c_1)}$ positiv sind, ist $\sum_{i=1}^n w_i^{(c_1)} > 0$. Sei nun $a^* := \frac{1}{(\sum_{i=1}^n w_i^{(c_1)})^2} > 0$ und w^* die eindeutige Optimallösung des obigen Problems mit $a = a^*$. Dann folgt

$$\sum_{i=1}^n w_i^* = \sum_{i=1}^n \sqrt{a^*} w_i^{(c_1)} = 1, \quad (3.10)$$

womit $w^* \in \mathcal{RB}_b(\mathbb{R}_{I_n,1}^n)$.

Die Eindeutigkeit dieser Lösung zeigen wir, indem wir beweisen, dass w^* unabhängig von der Wahl des Parameters c ist:

Dafür wenden wir nun obige Überlegungen auf $c = c_2 > 0$ an. Dann ist auch $w^{**} := \sqrt{a^{**}} w^{(c_2)} \in \mathcal{RB}_b(\mathbb{R}_{I_n,1}^n)$, wobei $w^{(c_2)}$ das eindeutige Minimum von (LS^{c_2}) ist und $a^{**} := \frac{1}{(\sum_{i=1}^n w_i^{(c_2)})^2} > 0$. $w^{(c_2)}$ ist jedoch auch die eindeutige Lösung von $(LS^{\bar{a}c_1})$ mit $\bar{a} = c_2/c_1 > 0$, womit $w^{(c_2)} = \sqrt{\bar{a}} w^{(c_1)}$ folgt. Wegen der normierten Gewichte von w^{**} gilt

$$\sum_{i=1}^n w_i^{**} = \sum_{i=1}^n \sqrt{a^{**}} w_i^{(c_2)} = \sum_{i=1}^n \sqrt{a^{**}} \sqrt{\bar{a}} w_i^{(c_1)} = 1. \quad (3.11)$$

Zusammen mit (3.10) erhalten wir $\sqrt{a^{**}} \sqrt{\bar{a}} = \sqrt{a^*}$ und insgesamt

$$w^{**} = \sqrt{a^{**}} \sqrt{\bar{a}} w^{(c_1)} = \sqrt{a^*} w^{(c_1)} = w^*.$$

□

Als nächstes möchten wir die Einschränkung auf „Long-Only“-Portfolios fallenlassen. Um das Optimierungsproblem (LS^c) verwenden zu können, müssen wir jedoch sicherstellen, dass der Logarithmus nur auf nichtnegative Werte angewendet wird. Dazu gehen wir wie Bai, Scheinberg und Tütüncü (2013) folgendermaßen vor:

3. Risikobasierte Portfoliooptimierung

Seien $T := \{-1, 1\}^n$, $\theta = (\theta_1, \dots, \theta_n)^\top \in T$ und $\Theta := \text{diag}(\theta_1, \dots, \theta_n)$.

Wir betrachten nun die Zulässigkeitsmenge $\mathbb{R}_\Theta^n = \{w \in \mathbb{R}^n \mid \Theta w \geq 0\}$. In der Menge der zulässigen Portfolios sind also von vornherein die Vorzeichen der einzelnen Assetgewichte festgelegt. Falls $\theta_i = 1$, so muss für das i -te Asset eine „Long Position“ eingenommen werden. Für $\theta_i = -1$ ist hingegen nur „Short Selling“ erlaubt. Wir möchten wieder ein „Risk-Budgeting“-Portfolio finden, sodass alle Assetgewichte ungleich 0 sind (vgl. Lemma 3.3.1). Dazu betrachten wir folgende Modifikation des Optimierungsproblems (LS^c):

$$\begin{aligned} (\text{LS}_\theta^c) \quad & \min_{w \in \mathbb{R}^n} \frac{1}{2} w^\top \Sigma w - c \sum_{i=1}^n b_i \ln \theta_i w_i \\ & \text{s.t.} \quad \theta_i w_i \geq 0, \quad i = 1, \dots, n \end{aligned}$$

Die Zielfunktion dieses Problems (als Abbildung von \mathbb{R}_Θ^n nach $\overline{\mathbb{R}}$) ist wieder für jedes beliebige $\theta \in T$ strikt konvex. Wir erhalten folgende Aussage:

Satz 3.3.5. *Sei $\Sigma \succ 0$ und $b \in \mathbb{R}_{+,1}^n$. Dann existiert für jedes beliebige $c > 0$ genau ein „Risk-Budgeting“-Portfolio $w^{(c)} \in \mathbb{R}_\Theta^n$ bezüglich b mit Risikobeiträgen $\zeta_i^2(w^{(c)}) = b_i c$, $\forall i = 1, \dots, n$, wobei dieses $w^{(c)}$ das eindeutige Minimum von (LS _{θ} ^c) ist und $\Theta w^{(c)} \in \mathbb{R}_+^n$ gilt.*

Beweis. Analog zu Lemma 3.3.2 und Satz 3.3.3 unter Verwendung von \mathbb{R}_Θ^n statt $\mathbb{R}_{I_n}^n$ und (LS _{θ} ^c) statt (LS^c). \square

Inkludieren wir nun wieder die Normierungsrestriktion $\sum_{i=1}^n w_i = 1$ in der Zulässigkeitsmenge, so lässt sich folgendes Korollar über die Existenz und Eindeutigkeit einer entsprechenden „Risk-Budgeting“-Lösung formulieren (vgl. Bai, Scheinberg und Tütüncü (2013) für den „Risk-Parity“-Spezialfall):

Korollar 3.3.6. *Sei $\Sigma \succ 0$ und $b \in \mathbb{R}_{+,1}^n$. Dann existiert höchstens ein „Risk-Budgeting“-Portfolio $w^* \in \{\mathbb{R}_{\Theta,1}^n \cup \mathbb{R}_{-\Theta,1}^n\}$ bezüglich b .*

Beweis. Lösen wir für ein beliebiges $c > 0$ die Optimierungsprobleme (LS _{θ} ^c) und (LS _{$-\theta$} ^c), so erhalten wir eindeutige Minima $w_\theta^{(c)}$ bzw. $w_{-\theta}^{(c)}$. Mit $v_i := -w_i$ lässt sich (LS _{$-\theta$} ^c) schreiben als

$$\begin{aligned} \min_{v \in \mathbb{R}^n} \quad & \frac{1}{2} v^\top \Sigma v - c \sum_{i=1}^n b_i \ln \theta_i v_i \\ & \text{s.t.} \quad \theta_i v_i \geq 0, \quad i = 1, \dots, n \end{aligned}$$

weshalb $w_\theta^{(c)} = -w_{-\theta}^{(c)}$ gilt. Wir unterscheiden nun 3 Fälle:

3. Risikobasierte Portfoliooptimierung

1. Fall: $e^\top w_\theta^{(c)} > 0 \Leftrightarrow e^\top w_{-\theta}^{(c)} < 0$:

Dann lassen sich alle Schritte im Beweis von Korollar 3.3.4, welches den Spezialfall $\theta = (1, \dots, 1)^\top$ behandelt, analog mit (LS_θ^c) durchführen. Es existiert also genau ein Portfolio $w^* \in \mathcal{RB}_b(\mathbb{R}_{\theta,1}^n)$, nämlich $w^* = \frac{w_\theta^{(c)}}{e^\top w_\theta^{(c)}}$. Da $\frac{w_{-\theta}^{(c)}}{e^\top w_{-\theta}^{(c)}} = \frac{w_\theta^{(c)}}{e^\top w_\theta^{(c)}} = w^*$, ist $\mathcal{RB}_b(\mathbb{R}_{-\theta,1}^n) = \emptyset$.

2. Fall: $e^\top w_\theta^{(c)} < 0 \Leftrightarrow e^\top w_{-\theta}^{(c)} > 0$:

Analog zum 1. Fall ist nun $w^* = \frac{w_{-\theta}^{(c)}}{e^\top w_{-\theta}^{(c)}}$ das einzige Element in $\mathcal{RB}_b(\mathbb{R}_{-\theta,1}^n)$ und $\mathcal{RB}_b(\mathbb{R}_{\theta,1}^n) = \emptyset$.

3. Fall: $e^\top w_\theta^{(c)} = e^\top w_{-\theta}^{(c)} = 0$:

Falls sich die Assetgewichte von $w_\theta^{(c)}$ zu 0 aufsummieren, so gilt dies auch für $w_{-\theta}^{(c)}$. Diese beiden Portfolios heißen dann auch *marktneutral*. Offensichtlich gibt es dann keine „Risk-Budgeting“-Lösung bezüglich b in den entsprechenden Zulässigkeitsmengen, da eine Gewichtsnormierung auf 1 nicht möglich ist.

□

Damit wissen wir also, dass nicht zwangsläufig ein „Risk-Budgeting“-Portfolio existieren muss, in dem der Investor sowohl „Long Positions“ als auch „Short Positions“ einnehmen kann. Aus den vorigen beiden Korollaren können wir nun jedoch folgende bedeutende Aussage ableiten:

Korollar 3.3.7. Sei $\Sigma \succ 0$ und $b \in \mathbb{R}_{+,1}^n$. Dann existieren in \mathbb{R}_1^n höchstens 2^{n-1} „Risk-Budgeting“-Portfolios bezüglich b . Es gilt also

$$1 \leq |\mathcal{RB}_b(\mathbb{R}_1^n)| \leq 2^{n-1}.$$

Beweis.

- Die untere Schranke folgt unmittelbar aus Korollar 3.3.4, weil

$$\mathcal{RB}_b(\mathbb{R}_1^n) \supseteq \mathcal{RB}_b(\mathbb{R}_{I_n,1}^n) = 1.$$

- Es gilt $\theta \in \{-1, 1\}^n =: T$ mit $|T| = 2^n$. Nach Korollar 3.3.6 gibt es zu jedem Paar $\{\theta, -\theta\}$ höchstens eine „Risk-Budgeting“-Lösung bezüglich b . Da

$$\mathcal{RB}_b(\mathbb{R}_1^n) = \bigsqcup_{\theta \in T} \mathcal{RB}_b(\mathbb{R}_{\theta,1}^n),$$

folgt $|\mathcal{RB}_b(\mathbb{R}_1^n)| \leq 2^n/2 = 2^{n-1}$.

□

3. Risikobasierte Portfoliooptimierung

Es gibt also nur eine endliche Anzahl von „Risk-Budgeting“-Lösungen, wenn die Menge der zulässigen Portfolios durch eine Normierung der Gewichte eingeschränkt wird. Möchte man ein beliebiges solches Portfolio ohne Vorgaben an die Vorzeichen der Assetgewichte mithilfe der oben definierten Optimierungsprobleme mit logarithmischer Schranke berechnen, so bietet es sich an, sich trotzdem auf den „Long-Only“-Fall zu beschränken, da dort die Existenz garantiert ist. Setzt man sich jedoch beispielsweise zum Ziel, das Portfolio mit der geringsten Portfolio-Varianz in $\mathcal{RB}_b(\mathbb{R}_+^n)$ zu finden, so ist die Anzahl der zu bestimmenden Portfolios exponentiell in der Anzahl der Assets. Unter Verwendung der obigen Überlegungen muss nämlich das Optimierungsproblem (LS_θ^c) für 2^{n-1} verschiedene Werte von θ in enumerativer Art und Weise gelöst werden. Auch für eine Zulässigkeitsmenge mit weiteren Restriktionen zur Einschränkung der einzelnen Assetgewichte, die eventuell zur Unzulässigkeit des „Risk-Budgeting“-Problems führen können, müssen diese 2^{n-1} Optimierungsprobleme untersucht werden, um überhaupt feststellen zu können, ob Zulässigkeit vorliegt oder nicht.

3.3.1.1. Das „Risk-Budgeting“-Problem im Vergleich zu anderen Ansätzen

Als Nächstes möchten wir die Varianz einer „Risk-Budgeting“-Lösung mit den Varianzen der Lösungen zweier anderer Ansätze vergleichen. Dabei wird wieder das Optimierungsproblem (LS^c) hilfreich sein.

Wir haben bereits das „Minimum-Variance“-Modell kennengelernt, welches die Varianz minimiert und damit - wie wir im praktischen Teil dieser Arbeit noch anhand einiger empirischer Beispiele sehen werden - eine starke Konzentration der Gewichte auf die risikoarmen Assets legt. Einerseits ist zwar eine Varianzminimierung im Sinne eines jeden Investors, auf der anderen Seite verringert eine starke Fokussierung auf Assets mit geringem Risiko die Diversifikation.

Wir betrachten zur Illustration folgendes Beispiel: Ein Investor möchte in zwei Aktien investieren und entscheidet sich für ein „Minimum-Variance“-Portfolio, wobei die Schätzung der Kovarianzmatrix Σ anhand historischer Daten durchgeführt wird. Diese Schätzung ergibt für die erste Aktie eine sehr hohe und für die zweite Aktie eine vergleichsweise niedrige Varianz, weshalb der „Minimum-Variance“-Ansatz zu einer hohen Gewichtung der zweiten Aktie führt. Stellt sich jedoch heraus, dass der historische Aktienverlauf der zweiten Aktie mit wenig Volatilität für die Zukunft nicht aussagekräftig und damit die durchgeführte Schätzung inadäquat ist (beispielsweise aufgrund einer plötzlich eintretenden Wirtschaftskrise), könnte die zu starke Konzentration auf diese Aktie Nachteile mit sich bringen.

Eine gänzlich andere Portfoliokonstruktion liefert der „Weight-Budgeting“-Ansatz. Dabei werden die Assetgewichte folgendermaßen gewählt:

$$w_i = b_i, \forall i = 1, \dots, n \quad \Leftrightarrow \quad w_i/b_i = w_j/b_j, \forall i, j = 1, \dots, n$$

3. Risikobasierte Portfoliooptimierung

Dabei wird wieder $b_i > 0, \forall i = 1, \dots, n$ und $\sum_{i=1}^n b_i = 1$ vorausgesetzt. Durch diese Konstruktion resultiert also immer ein „Long-Only“-Portfolio. Mit $b_i = 1/n, \forall i = 1, \dots, n$ werden beispielsweise die Gewichte gleichmäßig auf alle Assets verteilt, womit eine unerwartet hohe Volatilität eines einzelnen (a priori als risikoarm eingeschätzten) Assets keine so großen Auswirkungen hat wie im „Minimum-Variance“-Fall. Allerdings bleibt in diesem Modell das Risiko der Assets und damit auch die im Allgemeinen vorausgesetzte Risikoaversität eines Investors gänzlich unberücksichtigt. „Minimum-Variance“ und „Weight-Budgeting“ können somit als stark konträre Herangehensweisen angesehen werden.

Wir vergleichen nun folgende drei Modelle unter Verwendung der Zulässigkeitsmenge $\mathbb{R}_{I_n,1}^n = \{w \in \mathbb{R}^n \mid I_n w \geq 0, e^\top w = 1\}$ (siehe Maillard, Roncalli und Teiletche (2010) sowie Roncalli (2013)):

$$\begin{aligned} \frac{w_i}{b_i} &= \frac{w_j}{b_j}, \quad \forall i, j && \text{ („Weight-Budgeting“)} \\ \partial_{w_i} \varsigma^2(w) &= \partial_{w_j} \varsigma^2(w), \quad \forall i, j \text{ mit } w_i, w_j > 0 && \text{ („Minimum-Variance“)} \\ \frac{w_i \cdot \partial_{w_i} \varsigma^2(w)}{b_i} &= \frac{w_j \cdot \partial_{w_j} \varsigma^2(w)}{b_j}, \quad \forall i, j && \text{ („Risk-Budgeting“)} \end{aligned}$$

Unter Berücksichtigung dieser Bedingungen lässt sich das „Risk-Budgeting“-Modell als Kompromiss zwischen den beiden anderen Modellen interpretieren. Dieser Eindruck wird nun noch durch folgende Eigenschaft untermauert:

Wir untersuchen eine konkrete Instanz der obigen drei Probleme mit $\Sigma \succ 0$ und $b \in \mathbb{R}_{+,1}^n$. Mit $w^{(\text{WB})}$, $w^{(\text{MV})}$ und $w^{(\text{RB})}$ bezeichnen wir die entsprechenden Lösungen.⁶ Dann gilt folgende Aussage:

$$w^{(\text{MV})\top} \Sigma w^{(\text{MV})} \stackrel{(1)}{\leq} w^{(\text{RB})\top} \Sigma w^{(\text{RB})} \stackrel{(2)}{\leq} w^{(\text{WB})\top} \Sigma w^{(\text{WB})} \quad (3.12)$$

Das Risiko des „Risk-Budgeting“-Portfolios liegt also genau zwischen den Risiken des „Minimum-Variance“- und des „Weight-Budgeting“-Portfolios.

Ungleichung (1) in (3.12) folgt direkt aus der Tatsache, dass die Varianz von $w^{(\text{MV})}$ minimal unter allen Portfolios aus $\mathbb{R}_{I_n,1}^n$ ist. Ungleichung (2) lässt sich folgendermaßen zeigen:

Wir wissen bereits, dass $w^{(\text{RB})}$ die eindeutige Lösung des Optimierungsprogramms (LS^c) für ein spezielles $c = c^* > 0$ ist. Andererseits ist auch $w^{(\text{WB})}$ ein zulässiges Portfolio für dieses Problem. Wegen der Optimalität von $w^{(\text{RB})}$ gilt daher:

$$\frac{1}{2} w^{(\text{RB})\top} \Sigma w^{(\text{RB})} - c^* \sum_{i=1}^n b_i \ln w_i^{(\text{RB})} \leq \frac{1}{2} w^{(\text{WB})\top} \Sigma w^{(\text{WB})} - c^* \sum_{i=1}^n b_i \ln w_i^{(\text{WB})} \quad (3.13)$$

⁶Alle drei Portfolios $w^{(\text{WB})}$, $w^{(\text{MV})}$ und $w^{(\text{RB})}$ existieren und sind eindeutig bestimmt. Dies ist für $w^{(\text{WB})}$ trivial und folgt für $w^{(\text{MV})}$ aus der positiven Definitheit der Kovarianzmatrix bzw. für $w^{(\text{RB})}$ aus Korollar 3.3.4.

3. Risikobasierte Portfoliooptimierung

Wir lösen nun das folgende konvexe Optimierungsproblem mit strikt konvexer Zielfunktion:

$$\begin{aligned} \min_{w \in \mathbb{R}^n} \quad & - \sum_{i=1}^n b_i \ln w_i \\ \text{s.t.} \quad & e^\top w = 1 \end{aligned}$$

Aus den KKT-Bedingungen $e^\top w = 1$ und $-b_i/w_i - \gamma = 0, \forall i = 1, \dots, n$ erhalten wir die eindeutige Optimallösung $w_i = b_i = w_i^{(\text{WB})}$. Daher folgt wegen $c^* > 0$

$$-c^* \sum_{i=1}^n b_i \ln w_i^{(\text{RB})} \geq -c^* \sum_{i=1}^n b_i \ln w_i^{(\text{WB})}.$$

Gemeinsam mit (3.13) erhalten wir somit Ungleichung (2) in (3.12).

3.3.2. Eine Kleinste-Quadrate-Formulierung

In diesem Teil der Arbeit möchten wir ein weiteres Optimierungsproblem betrachten, das zum Auffinden von „Risk-Budgeting“-Portfolios verwendet werden kann und im Zentrum unserer weiteren theoretischen und praktischen Untersuchungen stehen wird. Dieses nichtlineare Programm, welches die quadrierten Abweichungen der einzelnen Asset-Risikobeiträge vom „Risk-Budgeting“-Zustand in der Zielfunktion berücksichtigt, wurde in Maillard, Roncalli und Teiletche (2010) für das „Risk-Parity“-Problem (in der von uns definierten Form) und in Roncalli (2013) für das „Risk-Budgeting“-Problem als Lösungsansatz vorgeschlagen. Bai, Scheinberg und Tütüncü (2013) haben in ihrem Paper *„Least-squares approach to risk parity in portfolio selection“* verschiedene Eigenschaften des Problems für den „Risk-Parity“-Fall genauer untersucht. Darauf basieren die folgenden Ausführungen, wobei wir jedoch den allgemeineren „Risk-Budgeting“-Ansatz zugrunde legen.

Im letzten Abschnitt hat uns ein konvexes Optimierungsproblem mit einer logarithmischen Schranke in der (strikt konvexen) Zielfunktion wertvolle Informationen über die Existenz und Eindeutigkeit von „Risk-Budgeting“-Lösungen für verschiedene Zulässigkeitsmengen geliefert. Dort haben wir allerdings nur Bedingungen an die Vorzeichen und die Normierung der Assetgewichte gestellt. Oftmals interessieren uns jedoch auch restriktivere Zulässigkeitsmengen der Form $\{w \in \mathbb{R}^n \mid L_i \leq w_i \leq U_i, \forall i = 1, \dots, n \text{ und } e^\top w = 1\}$ mit oberen und unteren Schranken für die einzelnen Gewichte. Wir werden für das folgende Kleinste-Quadrate-Programm eine noch etwas allgemeinere Menge verwenden, die an die Zulässigkeitsmengen der Markowitz’schen „Mean-Variance“-Probleme des 2. Kapitels angelehnt ist, nämlich

$$\Omega_{KQ} := \{w \in \mathbb{R}^n \mid Aw \geq \alpha \text{ und } Bw = \beta\}$$

mit $A \in \mathbb{R}^{p \times n}, \alpha \in \mathbb{R}^p, B \in \mathbb{R}^{q \times n}$ und $\beta \in \mathbb{R}^q$. In diese Menge lassen sich auch die Bedingungen $\mu^\top w \geq r$ oder $\mu^\top w = r$ an den erwarteten Portfolio-Return einbetten, was bei einer vergleichenden Analyse mit dem Markowitz’schen Modell hilfreich ist.

3. Risikobasierte Portfoliooptimierung

Unser Ziel ist nun wieder, ein nichtlineares Programm der Form

$$\begin{aligned} \min_{w \in \mathbb{R}^n} \quad & f(w) \\ \text{s.t.} \quad & Aw \geq \alpha \\ & Bw = \beta \end{aligned}$$

zu finden, sodass die Optimallösungen dieses Problems (unter gewissen Voraussetzungen) auch „Risk-Budgeting“-Portfolios sind. Darüber hinaus möchten wir jedoch auch erreichen, dass ein Minimum den „Risk-Budgeting“-Zustand sinnvoll approximiert, falls die Zulässigkeitsmenge kein „Risk-Budgeting“-Portfolio beinhaltet.⁷ In der Formulierung mit logarithmischer Schranke aus dem vorhergehenden Abschnitt könnten zwar prinzipiell auch beliebige zusätzliche Restriktionen eingebaut werden, doch wenn dies dazu führt, dass keine zulässige „Risk-Budgeting“-Lösung existiert, so stellt das Minimum von (LS_θ^c) keine zweckmäßige Approximation dar.

Eine mögliche Wahl für die Zielfunktion dieses Optimierungsproblems, welche Roncalli (2013) angibt, ist

$$f(w) = \sum_{i=1}^n \sum_{j=1}^n \left(\frac{w_i(\Sigma w)_i}{b_i} - \frac{w_j(\Sigma w)_j}{b_j} \right)^2 = \sum_{i=1}^n \sum_{j=1}^n \left(\frac{\zeta_i^2(w)}{b_i} - \frac{\zeta_j^2(w)}{b_j} \right)^2.$$

Diese Funktion summiert also über die quadrierten Abweichungen der einzelnen, durch die entsprechenden (vorgegebene) Risikoanteile dividierten Risikobeiträge voneinander und bildet offensichtlich von \mathbb{R}^n nach \mathbb{R}_0^+ ab. Jedes Portfolio w^* mit $f(w^*) = 0$ erfüllt $\frac{\zeta_i^2(w^*)}{b_i} = \frac{\zeta_j^2(w^*)}{b_j}$, $\forall i, j = 1, \dots, n$ und ist damit im „Risk-Budgeting“-Zustand bezüglich b .

Wir werden jedoch im Weiteren die beiden folgenden Optimierungsprobleme verwenden, deren Zielfunktionen den selben Zweck wie die obige Funktion erfüllen, jedoch nur aus $\mathcal{O}(n)$ (statt $\mathcal{O}(n^2)$) Summanden bestehen:

$$\begin{aligned} \text{(KQ1)} \quad \min_{w \in \mathbb{R}^n} \quad & \sum_{i=1}^n b_i \left(\frac{w_i(\Sigma w)_i}{b_i} - \sum_{j=1}^n w_j(\Sigma w)_j \right)^2 = \sum_{i=1}^n b_i \left(\frac{\zeta_i^2(w)}{b_i} - w^\top \Sigma w \right)^2 \\ \text{s.t.} \quad & Aw \geq \alpha \\ & Bw = \beta \end{aligned}$$

$$\begin{aligned} \text{(KQ2)} \quad \min_{\substack{w \in \mathbb{R}^n \\ c \in \mathbb{R}}} \quad & \sum_{i=1}^n b_i \left(\frac{w_i(\Sigma w)_i}{b_i} - c \right)^2 = \sum_{i=1}^n b_i \left(\frac{\zeta_i^2(w)}{b_i} - c \right)^2 \\ \text{s.t.} \quad & Aw \geq \alpha \\ & Bw = \beta \end{aligned}$$

⁷Dies ist beispielsweise schon der Fall, wenn die Return-Bedingung $\mu^\top w \geq r$ mit einer zu restriktiven Schranke r zur Anwendung kommt.

3. Risikobasierte Portfoliooptimierung

Die Zielfunktion von (KQ1) beinhaltet dabei für ein Portfolio w die mit den Risikoanteilen b_i gewichteten quadrierten Differenzen zwischen den tatsächlichen Risikobeiträgen und den gewünschten Risikobeiträgen im „Risk-Budgeting“-Zustand, jeweils skaliert mit den Risikoanteilen b_i . Im Programm (KQ2) wird hingegen der Wert $w^\top \Sigma w$ durch eine Variable c ersetzt, über die simultan mit dem Portfolio w ebenfalls minimiert wird. Diese Herangehensweise wird in ähnlicher Weise in Bai, Scheinberg und Tütüncü (2013) für das „Risk-Parity“-Problem gewählt. Wir werden im Folgenden die Notationen

- $f_1(w) := \sum_{i=1}^n b_i \left(\frac{w_i(\Sigma w)_i}{b_i} - w^\top \Sigma w \right)^2$
- $f_2(w, c) := \sum_{i=1}^n b_i \left(\frac{w_i(\Sigma w)_i}{b_i} - c \right)^2$

für die Zielfunktionen der beiden Probleme (KQ1) und (KQ2) verwenden. Folgendes Lemma zeigt den Zusammenhang zwischen diesen beiden Formulierungen:

Lemma 3.3.8. *Sei $b \in \mathbb{R}_{+,1}^n$ und $\hat{w} \in \Omega_{KQ}$ ein beliebiges, zulässiges Portfolio. Dann hat das Problem $\min_{c \in \mathbb{R}} \{f_2(w, c) \mid w = \hat{w}\}$ genau ein globales Optimum \hat{c} , nämlich $\hat{c} = \hat{w}^\top \Sigma \hat{w}$.*

Beweis. (Vgl. Bai, Scheinberg und Tütüncü, 2013.)

Für ein festes $w = \hat{w}$ ist die Funktion $f_2(\hat{w}, c)$ strikt konvex in $c \in \mathbb{R}$, da

$$\frac{\partial^2}{\partial c^2} f_2(\hat{w}, c) = 2 \cdot \sum_{i=1}^n b_i = 2 > 0, \forall c \in \mathbb{R}.$$

Daher ist \hat{c} das eindeutige Minimum, falls $\frac{\partial}{\partial c} f_2(\hat{w}, c) \Big|_{c=\hat{c}} = 0$ gilt. Somit folgt:

$$-2 \cdot \sum_{i=1}^n b_i \left(\frac{\hat{w}_i(\Sigma \hat{w})_i}{b_i} - \hat{c} \right) = 0 \quad \Rightarrow \quad \hat{c} = \frac{\sum_{i=1}^n \hat{w}_i(\Sigma \hat{w})_i}{\sum_{i=1}^n b_i} = \hat{w}^\top \Sigma \hat{w}$$

□

Wie wir sehen werden, ist es trotz dieser Eigenschaft manchmal sinnvoll, die Problemformulierung (KQ2) mit einem variablen Parameter c zu verwenden.

Wir möchten nun noch kurz auf die Zulässigkeitsmengen $\mathbb{R}_\Theta^n = \{w \in \mathbb{R}^n \mid \Theta w \geq 0\}$ sowie $\mathbb{R}_{\Theta,1}^n = \{w \in \mathbb{R}^n \mid \Theta w \geq 0, e^\top w = 1\}$ mit $\theta \in T = \{-1, 1\}^n$ und $\Theta = \text{diag}(\theta_1, \dots, \theta_n)$ aus Kapitel 3.3.1 eingehen:

Wie wir bereits wissen, gibt es (mit $b \in \mathbb{R}_{+,1}^n$) für jedes beliebige $c > 0$ ein Portfolio $w^{(c)} \in \mathcal{RB}_b(\mathbb{R}_\Theta^n)$ mit Risikobeiträgen $\zeta_i^2(w^{(c)}) = b_i c, \forall i = 1, \dots, n$. Daraus lässt sich (falls die Summe der Assetgewichte ungleich 0 ist) durch Normierung ein „Risk-Budgeting“-Portfolio berechnen, das in $\{\mathbb{R}_{\Theta,1}^n \cup \mathbb{R}_{-\Theta,1}^n\}$ liegt. Deswegen besteht eine Möglichkeit zur Berechnung eines Portfolios $w^* \in \mathcal{RB}_b(\{\mathbb{R}_{\Theta,1}^n \cup \mathbb{R}_{-\Theta,1}^n\})$ auch darin, die Funktion $f_2(w, c)$ mit einem beliebigen, festen $c = \hat{c} > 0$ über alle $w \in \mathbb{R}_\Theta^n$ zu minimieren und erst anschließend die Gewichtsnormierung durchzuführen.

3. Risikobasierte Portfoliooptimierung

Im Gegensatz zu den Problemen (LS^c) bzw. (LS_θ^c) sind die Zielfunktionen von (KQ1) und (KQ2) nicht konvex, womit keine konvexen Optimierungsprobleme (KP) vorliegen. Die KKT-Bedingungen 1. Ordnung sind daher für diese beiden Probleme nicht zwangsläufig ausreichend, um eine Optimallösung bestimmen zu können. Wie wir nun sehen werden, ist jedoch unter gewissen zusätzlichen Voraussetzungen jeder KKT-Punkt trotzdem ein globales Minimum. Dafür formulieren wir die KKT-Bedingungen für (KQ2):

Sei

$$\mathcal{L}_2(w, c, \lambda, \gamma) := \sum_{i=1}^n b_i \left(\frac{w_i(\Sigma w)_i}{b_i} - c \right)^2 - \lambda^\top (Aw - \alpha) - \gamma^\top (Bw - \beta)$$

die Lagrangefunktion von (KQ2). Ein Vektor $(w^*, c^*, \lambda, \gamma) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q$ ist ein KKT-Punkt für (KQ2), falls die folgenden Bedingungen erfüllt sind:

- [2-i] $Aw^* \geq \alpha$, $Bw^* = \beta$, d.h. w^* ist zulässig für (KQ2),
- [2-ii] $\nabla_w \mathcal{L}_2(w, c, \lambda, \gamma) \Big|_{\substack{w=w^* \\ c=c^*}} = \nabla_w f_2(w, c) \Big|_{\substack{w=w^* \\ c=c^*}} - A^\top \lambda - B^\top \gamma = 0$,
- [2-iii] $\frac{\partial}{\partial c} \mathcal{L}_2(w, c, \lambda, \gamma) \Big|_{\substack{w=w^* \\ c=c^*}} = \frac{\partial}{\partial c} f_2(w, c) \Big|_{\substack{w=w^* \\ c=c^*}} = 0$,
- [2-iv] $\lambda \geq 0$, $\lambda^\top \cdot (Aw^* - \alpha) = 0$.

Sei nun $e_i = (0, \dots, 0, 1, 0, \dots, 0)^\top \in \mathbb{R}^n$ die i -te Spalte der n -dimensionalen Einheitsmatrix I_n und $\Sigma_i = (\sigma_{1,i}, \dots, \sigma_{n,i})^\top = (\sigma_{i,1}, \dots, \sigma_{i,n})^\top \in \mathbb{R}^n$ die i -te Spalte von Σ . Damit gilt

$$\begin{aligned} \Sigma_i e_i^\top w &= (\sigma_{i,1} w_1, \dots, \sigma_{i,n} w_n)^\top, \\ e_i \Sigma_i^\top w &= \left(0, \dots, 0, \sum_{j=1}^n \sigma_{i,j} w_j, 0, \dots, 0 \right)^\top. \end{aligned}$$

Wir erhalten für die partiellen Ableitungen von f_2 nach w_k

$$\begin{aligned} \frac{\partial}{\partial w_k} f_2(w, c) &= 2 \cdot \sum_{\substack{i=1 \\ i \neq k}}^n b_i \left(\frac{w_i(\Sigma w)_i}{b_i} - c \right) \cdot \frac{\sigma_{i,k} w_i}{b_i} \\ &\quad + 2 \cdot b_k \left(\frac{w_k(\Sigma w)_k}{b_k} - c \right) \cdot \frac{1}{b_k} \left(\sum_{\substack{j=1 \\ j \neq k}}^n \sigma_{k,j} w_j + 2\sigma_{k,k} w_k \right) \\ &= 2 \cdot \sum_{i=1}^n \left(\frac{w_i(\Sigma w)_i}{b_i} - c \right) \cdot \sigma_{i,k} w_i + 2 \cdot \left(\frac{w_k(\Sigma w)_k}{b_k} - c \right) \cdot \left(\sum_{j=1}^n \sigma_{k,j} w_j \right) \\ &= 2 \cdot \sum_{i=1}^n \left(\frac{w_i(\Sigma w)_i}{b_i} - c \right) \cdot (\Sigma_i e_i^\top w)_k + 2 \cdot \left(\frac{w_k(\Sigma w)_k}{b_k} - c \right) \cdot (e_k \Sigma_k^\top w)_k. \end{aligned}$$

3. Risikobasierte Portfoliooptimierung

Der Gradient von f_2 bezüglich w lässt sich somit schreiben als

$$\begin{aligned}\nabla_w f_2(w, c) &= 2 \cdot \sum_{i=1}^n \left(\frac{w_i(\Sigma w)_i}{b_i} - c \right) \cdot (\Sigma_i e_i^\top w + e_i \Sigma_i^\top w) \\ &= 2 \cdot \sum_{i=1}^n \left(\frac{w_i(\Sigma w)_i}{b_i} - c \right) \cdot (\Sigma_i e_i^\top + e_i \Sigma_i^\top) w.\end{aligned}$$

Folgendes Lemma besagt nun, dass jeder KKT-Punkt von (KQ2) auch ein globales Minimum ist, falls dieses Problem ohne Nebenbedingungen gelöst wird:

Lemma 3.3.9. *Betrachte folgendes Optimierungsproblem als Spezialfall von (KQ2):*

$$\min_{w \in \mathbb{R}^n, c \in \mathbb{R}} f_2(w, c)$$

Dann ist jeder KKT-Punkt dieses Problems ein globales Optimum.

Beweis. (Siehe Bai, Scheinberg und Tütüncü (2013) für den „Risk-Parity“-Spezialfall.) Sei (w^*, c^*) ein KKT-Punkt. Dieser erfüllt folgende Bedingungen:

$$[2^*-i] \quad \nabla_w f_2(w, c) \Big|_{\substack{w=w^* \\ c=c^*}} = 0,$$

$$[2^*-ii] \quad \frac{\partial}{\partial c} f_2(w, c) \Big|_{\substack{w=w^* \\ c=c^*}} = 0.$$

Aus $[2^*-i]$ folgt, dass auch

$$\begin{aligned}w^{*\top} \cdot \nabla_w f_2(w, c) \Big|_{\substack{w=w^* \\ c=c^*}} &= 2 \cdot \sum_{i=1}^n \left(\frac{w_i^*(\Sigma w^*)_i}{b_i} - c^* \right) \cdot w^{*\top} (\Sigma_i e_i^\top + e_i \Sigma_i^\top) w^* \\ &= 2 \cdot \sum_{i=1}^n \left(\frac{w_i^*(\Sigma w^*)_i}{b_i} - c^* \right) \cdot (w_i^*(\Sigma w^*)_i + w_i^*(\Sigma w^*)_i) \\ &= 4 \cdot \sum_{i=1}^n \left(\frac{\varsigma_i^2(w^*)}{b_i} - c^* \right) \cdot \varsigma_i^2(w^*) \\ &= 0.\end{aligned}$$

Wie wir bereits im Beweis von Lemma 3.3.8 gesehen haben, folgt aus $[2^*-ii]$ unmittelbar $c^* = w^{*\top} \Sigma w^* = \sum_{i=1}^n \varsigma_i^2(w^*)$. Setzen wir dies in obige Gleichung ein, so erhalten wir die Bedingung

$$\sum_{i=1}^n \frac{1}{b_i} (\varsigma_i^2(w^*))^2 = \left(\sum_{i=1}^n \varsigma_i^2(w^*) \right)^2.$$

Wenden wir hingegen die Cauchy-Schwarz'sche Ungleichung mit den beiden Vektoren

$$\left(\frac{1}{\sqrt{b_1}} \varsigma_1^2(w^*), \dots, \frac{1}{\sqrt{b_n}} \varsigma_n^2(w^*) \right)^\top \in \mathbb{R}^n \quad \text{und} \quad (\sqrt{b_1}, \dots, \sqrt{b_n})^\top \in \mathbb{R}^n$$

3. Risikobasierte Portfoliooptimierung

an (siehe Appendix A.4), so wissen wir, dass

$$\begin{aligned} \left(\sum_{i=1}^n \sqrt{b_i} \right) \left(\sum_{i=1}^n \left(\frac{1}{\sqrt{b_i}} \varsigma_i^2(w^*) \right)^2 \right) &\geq \left(\sum_{i=1}^n \sqrt{b_i} \cdot \frac{1}{\sqrt{b_i}} \varsigma_i^2(w^*) \right)^2 \\ &\Leftrightarrow \\ \sum_{i=1}^n \frac{1}{b_i} (\varsigma_i^2(w^*))^2 &\geq \left(\sum_{i=1}^n \varsigma_i^2(w^*) \right)^2 \end{aligned}$$

gilt, wobei diese Ungleichung genau dann mit Gleichheit erfüllt ist, wenn die Vektoren linear abhängig sind.⁸ Dies bedeutet in unserem Fall, dass die Skalare $\frac{\varsigma_i^2(w^*)}{b_i}$ für alle $i = 1, \dots, n$ gleich einem konstanten Wert \tilde{c} sind. \tilde{c} ist nun jedoch genau c^* , da

$$\tilde{c} = \tilde{c} \cdot \sum_{i=1}^n b_i = \sum_{i=1}^n \varsigma_i^2(w^*) = c^*.$$

Daher folgt insgesamt für jeden beliebigen KKT-Punkt (w^*, c^*) , dass

$$f_2(w^*, c^*) = \sum_{i=1}^n b_i \left(\frac{\varsigma_i^2(w^*)}{b_i} - c^* \right)^2 = 0,$$

womit (w^*, c^*) global optimal ist. □

Sind wir also nur daran interessiert, ein beliebiges „Risk-Budgeting“-Portfolio (ohne Restriktionen) zu berechnen, stellt es keinen Nachteil dar, dass das Optimierungsproblem (KQ2) nicht konvex ist. Möchte man hingegen zusätzliche Restriktionen (wie beispielsweise eine untere Schranke für den erwarteten Return nach dem Schema des „Mean-Variance“-Optimierungsproblems (MW1) oder obere und untere Schranken für die Assetgewichte) verwenden, ermöglichen die Programme (KQ1) und (KQ2) überhaupt erst, diese sinnvoll einzubauen. Dort erfolgt nämlich eine Minimierung der Summe der quadrierten Abweichungen der Risikobeiträge vom „Risk-Budgeting“-Zustand. Damit sind Optima dieser Optimierungsprobleme auch für den Fall, dass in der Zulässigkeitsmenge gar keine „Risk-Budgeting“-Lösung liegt, adäquate Approximationen. Im Rahmen des praktischen Teils dieser Arbeit (Kapitel 5) werden wir dementsprechend die berechneten (eventuell auch nur lokalen) Minima von (KQ1) und (KQ2) für den Fall, dass sie nicht im „Risk-Budgeting“-Zustand sind, als *approximative „Risk-Budgeting“-Portfolios* bezeichnen.

⁸Wir nehmen hier o.B.d.A. implizit an, dass $w^* \neq 0$ und damit auch $\varsigma_i^2(w^*) \neq 0, \forall i = 1, \dots, n$. Im Fall $w^* = 0$ ist nämlich auch $c^* = 0$ und der korrespondierende KKT-Punkt (w^*, c^*) ist auf jeden Fall ein globales Optimum mit Zielfunktionswert 0.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

In diesem Kapitel möchten wir nun Algorithmen kennenlernen, mit denen es uns möglich sein wird, nichtlineare Optimierungsprobleme der Form (NLP) - also mit Gleichheits- und Ungleichheitsrestriktionen - zu lösen. Unser Hauptaugenmerk gilt dabei den Kleinst-Quadrat-Optimierungsproblemen (KQ1) bzw. (KQ2) aus Abschnitt 3.3.2 zur Berechnung von „Risk-Budgeting“-Portfolios.

Wir werden aus den unzähligen Methoden, die in den letzten Jahrzehnten entwickelt wurden, in erster Linie die Verfahren der sequentiellen quadratischen Programmierung (SQP) untersuchen. Diese versuchen, beliebige nichtlineare Programme mithilfe einer Sequenz von quadratischen Optimierungsproblemen der Form (QP) zu lösen. Daher beginnen wir mit speziellen Lösungsalgorithmen für quadratische Programme, welche wir jedoch nicht nur im Rahmen der SQP-Verfahren benötigen werden, sondern auch direkt auf die „Mean-Variance“-Optimierungsprobleme (MW1) und (MW3) aus Kapitel 2 anwenden und damit unmittelbar Portfoliooptimierung im Sinne von Markowitz durchführen können.

In **Abschnitt 4.1** über quadratische Optimierungsprobleme beschränken wir uns meist auf *konvexe* Programme. Diese Klasse von Optimierungsproblemen beinhaltet einerseits die für uns relevanten Markowitz’schen Problemformulierungen und ist andererseits auch für die von uns betrachteten SQP-Verfahren ausreichend. Daher konzentrieren wir uns (unter Berücksichtigung von Korollar 1.4.8) auf die Bestimmung von KKT-Punkten (vgl. Kapitel 1.4). Wir werden folgende Methoden genauer untersuchen:

- Lösung von quadratischen Problemen mit Gleichheitsrestriktionen
- Das Verfahren der aktiven Menge für quadratische Probleme mit Ungleichheitsrestriktionen (mit Konvergenzaussagen)
- Innere-Punkte-Verfahren für quadratische Probleme mit Ungleichheitsrestriktionen
 - Zulässige pfadverfolgende Innere-Punkte-Verfahren unter Verwendung von Nachbarschaften (mit Konvergenzaussagen)
 - Ein unzulässiges Innere-Punkte-Verfahren („Predictor-Corrector“-Verfahren nach Mehrotra)

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

Auch in **Abschnitt 4.2**, welcher die sequentielle quadratische Programmierung behandelt, versuchen wir, KKT-Punkte der betrachteten nichtlinearen Optimierungsprobleme zu finden. Da diese Probleme jedoch im Allgemeinen nicht konvex sind, werden zusätzliche Maßnahmen ergriffen, um beispielsweise lokale Maxima zu vermeiden. Wir werden nachstehende Verfahren kennenlernen:

- Die Lagrange-Newton-Methode für gleichheitsrestringierte Probleme
- Eine lokale SQP-Methode für Probleme mit Ungleichheitsrestriktionen (mit lokalen Konvergenzaussagen)
- Eine zulässige globale SQP-Methode für Probleme mit *linearen* Gleichheits- und Ungleichheitsrestriktionen (mit globalen Konvergenzaussagen)

Im Zusammenhang mit der Berechnung von KKT-Punkten werden wir mehrmals die Jacobi-Matrix bestimmter Funktionen benötigen. Für eine stetig differenzierbare Funktion $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ mit

$$F(x) = (f_1(x), \dots, f_m(x))^T$$

und $f_i : \mathbb{R}^n \rightarrow \mathbb{R}, \forall i = 1, \dots, m$ bezeichnen wir in den folgenden Kapiteln mit

$$\mathcal{J}_F(x) = \mathcal{J}_F(x_1, \dots, x_n) := \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

die Jacobi-Matrix von F im Punkt x .

4.1. Quadratische Optimierungsprobleme

4.1.1. Behandlung von Problemen mit Gleichheitsrestriktionen

Die Resultate dieses Abschnittes basieren auf Geiger und Kanzow (2002).

Wir betrachten zu Beginn einen Spezialfall des Optimierungsproblems (QP) aus Abschnitt 1.3, in dem die zulässige Menge ausschließlich durch Gleichheitsrestriktionen beschrieben wird. Wir suchen also eine Lösung des Programms

$$\begin{aligned} (\overline{\text{QP}}) \quad & \min_{x \in \mathbb{R}^n} \quad \frac{1}{2} x^T Q x + c^T x \\ & \text{s.t.} \quad Bx = \beta \end{aligned}$$

mit $Q \in \mathbb{R}^{n \times n}$ (o.B.d.A. symmetrisch), $c \in \mathbb{R}^n$, $B \in \mathbb{R}^{q \times n}$ und $\beta \in \mathbb{R}^q$.

Wenden wir nun Satz 1.4.1 in Verbindung mit Korollar 1.4.2 an, so wissen wir, dass für jedes lokale Minimum x^* von $(\overline{\text{QP}})$ ein Vektor $\gamma \in \mathbb{R}^q$ existiert, sodass (x^*, γ) ein KKT-Punkt ist. Die entsprechenden KKT-Bedingungen

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

- $Qx^* + c - B^\top \gamma = 0$,
- $Bx^* = \beta$,

können folgendermaßen als lineares Gleichungssystem geschrieben werden:

$$\begin{pmatrix} Q & -B^\top \\ B & 0 \end{pmatrix} \begin{pmatrix} x^* \\ \gamma \end{pmatrix} = \begin{pmatrix} -c \\ \beta \end{pmatrix} \quad (4.1)$$

Wie wir bereits wissen, ist $(\overline{\text{QP}})$ ein konvexes Optimierungsproblem, falls Q positiv semidefinit ist. Für diesen Fall stimmt die Menge aller Minima von $(\overline{\text{QP}})$ mit der Lösungsmenge des linearen Gleichungssystems (4.1) überein.

Im nächsten Abschnitt möchten wir diese Vorgangsweise auf allgemeine quadratische Programme, welche auch Ungleichheitsrestriktionen beinhalten, erweitern. Dafür schreiben wir (4.1) wie folgt um:

Sei (x^*, γ) ein KKT-Punkt, \hat{x} ein beliebiger *zulässiger* Punkt für $(\overline{\text{QP}})$ und $x^* = \hat{x} + \Delta\hat{x}$. Dann gilt:

$$\begin{aligned} \begin{pmatrix} Q & -B^\top \\ B & 0 \end{pmatrix} \begin{pmatrix} x^* \\ \gamma \end{pmatrix} &= \begin{pmatrix} -c \\ \beta \end{pmatrix} \\ \Leftrightarrow \\ \begin{pmatrix} Q & -B^\top \\ B & 0 \end{pmatrix} \begin{pmatrix} \Delta\hat{x} \\ \gamma \end{pmatrix} &= \begin{pmatrix} -c \\ \beta \end{pmatrix} - \begin{pmatrix} Q & -B^\top \\ B & 0 \end{pmatrix} \begin{pmatrix} \hat{x} \\ 0 \end{pmatrix} = \begin{pmatrix} -Q\hat{x} - c \\ \beta - B\hat{x} \end{pmatrix} = \begin{pmatrix} -Q\hat{x} - c \\ 0 \end{pmatrix} \end{aligned}$$

Dabei ist $-Q\hat{x} - c$ genau der negative Gradient der Zielfunktion von $(\overline{\text{QP}})$ an der Stelle \hat{x} .

Aus dieser Darstellung von (4.1) sowie Korollar 1.4.8 folgt nun unmittelbar der nachstehende Satz (vgl. Geiger und Kanzow, 2002):

Satz 4.1.1. *Für einen beliebigen zulässigen Vektor $\hat{x} \in \mathbb{R}^n$ des Optimierungsproblems $(\overline{\text{QP}})$ ist der Punkt $(x^*, \gamma) \in \mathbb{R}^n \times \mathbb{R}^q$ ein KKT-Punkt, genau dann wenn*

$$x^* = \hat{x} + \Delta\hat{x}$$

sowie

$$\begin{pmatrix} Q & -B^\top \\ B & 0 \end{pmatrix} \begin{pmatrix} \Delta\hat{x} \\ \gamma \end{pmatrix} = \begin{pmatrix} -Q\hat{x} - c \\ 0 \end{pmatrix}$$

erfüllt sind.

Ist zusätzlich zu den obigen Bedingungen die Matrix Q positiv semidefinit (bzw. positiv definit), so ist x^ ein globales (bzw. strikt globales) Minimum von $(\overline{\text{QP}})$.*

Auf Möglichkeiten zur effizienten Lösung dieses Gleichungssystems möchten wir im Rahmen dieser Arbeit nicht näher eingehen. Ein entsprechender Algorithmus unter Verwendung der QR-Zerlegung wird in Spellucci (1993) beschrieben. Weitere geeignete Lösungsverfahren sind in Nocedal und Wright (2006) sowie Alt (2010) zu finden.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

4.1.2. Die Methode der aktiven Menge für konvexe Probleme mit Ungleichheitsrestriktionen

Wir interessieren uns nun für Lösungen des quadratischen Problems (QP), wie wir es bereits in Abschnitt 1.3 definiert haben:

$$\begin{aligned} \text{(QP)} \quad & \min_{x \in \mathbb{R}^n} \quad \frac{1}{2} x^\top Q x + c^\top x \\ & \text{s.t.} \quad Ax \geq \alpha \\ & \quad \quad Bx = \beta \end{aligned}$$

mit $Q \in \mathbb{R}^{n \times n}$ (o.B.d.A. symmetrisch), $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{p \times n}$, $\alpha \in \mathbb{R}^p$, $B \in \mathbb{R}^{q \times n}$ und $\beta \in \mathbb{R}^q$, wobei wir die Notation

$$f(x) := \frac{1}{2} x^\top Q x + c^\top x$$

für die Zielfunktion verwenden werden.

Ein KKT-Punkt $(x^*, \lambda, \gamma) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q$ dieses Optimierungsproblems erfüllt folgende Voraussetzungen:

$$\begin{aligned} \text{[QP-i]} \quad & Ax^* \geq \alpha, \quad Bx^* = \beta, \\ \text{[QP-ii]} \quad & Qx^* + c - A^\top \lambda - B^\top \gamma = 0, \\ \text{[QP-iii]} \quad & \lambda \geq 0, \quad \lambda^\top \cdot (Ax^* - \alpha) = 0. \end{aligned}$$

Für jedes lokale Minimum x^* von (QP) existieren Vektoren λ und γ , sodass (x^*, λ, γ) ein KKT-Punkt ist (Satz 1.4.1 und Korollar 1.4.2). Ist Q positiv semidefinit (bzw. positiv definit), so sind die KKT-Bedingungen 1. Ordnung wiederum hinreichend für die globale (bzw. strikt globale) Optimalität (Korollar 1.4.8).

Die erste Herangehensweise, die wir genauer untersuchen möchten, ist die *Methode der aktiven Menge*. Wir werden uns dabei (wie bereits in der Einleitung zu Kapitel 4 erwähnt) auf konvexe quadratische Programme ($Q \succcurlyeq 0$) beschränken.

Sofern nicht anders angegeben, stammen die Ergebnisse dieses Kapitels aus Geiger und Kanzow (2002). Eine mögliche Modifikation dieses Verfahrens zur Behandlung des indefiniten quadratischen Optimierungsproblems ist in Spellucci (1993) nachzulesen.

Die grundlegende Idee ist hierbei, iterativ eine Folge von Optimierungsproblemen der Form $(\overline{\text{QP}})$ - also ohne Ungleichheitsrestriktionen - zu generieren, für welche (wie wir im letzten Kapitel gesehen haben) die Berechnung eines KKT-Punktes äquivalent zur Lösung eines linearen Gleichungssystems ist. Dabei wird in einem *zulässigen* Punkt x^0 von (QP) gestartet und in der $(K + 1)$ -ten Iteration jenes Problem $(\overline{\text{QP}})$ gelöst, dessen Zulässigkeitsmenge durch eine Approximation der Menge der in x^K *aktiven* Restriktionen beschrieben wird. Wir betrachten dieses gleichheitsrestringierte Optimierungsproblem formal etwas genauer,

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

wobei wir in weiterer Folge für die Matrizen A und B sowie die Vektoren α und β die Schreibweise

$$A = \begin{pmatrix} a_1^\top \\ a_2^\top \\ \vdots \\ a_p^\top \end{pmatrix} \quad B = \begin{pmatrix} b_1^\top \\ b_2^\top \\ \vdots \\ b_q^\top \end{pmatrix} \quad \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{pmatrix} \quad (4.2)$$

mit $a_i \in \mathbb{R}^n$, $\forall i = 1, \dots, p$ und $b_k \in \mathbb{R}^n$, $\forall k = 1, \dots, q$ verwenden werden. Darüber hinaus benötigen wir für eine Indexmenge $\mathcal{I} \subseteq \{1, \dots, p\}$ noch folgende Notationen:

$$A_{\mathcal{I}} := \begin{pmatrix} \vdots \\ a_i^\top (i \in \mathcal{I}) \\ \vdots \end{pmatrix} \in \mathbb{R}^{|\mathcal{I}| \times n} \quad \alpha_{\mathcal{I}} := \begin{pmatrix} \vdots \\ \alpha_i (i \in \mathcal{I}) \\ \vdots \end{pmatrix} \in \mathbb{R}^{|\mathcal{I}|} \quad (4.3)$$

Analog zu $\alpha_{\mathcal{I}}$ definieren wir $\lambda_{\mathcal{I}} := (\dots, \lambda_i (i \in \mathcal{I}), \dots)^\top$ für einen Vektor $\lambda \in \mathbb{R}^p$ von Lagrange-Multiplikatoren.

Die Methode der aktiven Menge arbeitet nun im $(K + 1)$ -ten Iterationsschritt mit einem zulässigen Punkt x^K und einer Approximation \mathcal{J}_K der Menge

$$J(x^K) = \{i \in \{1, \dots, p\} \mid a_i^\top x^K = \alpha_i\},$$

welche die Indizes der in x^K aktiven Ungleichheitsrestriktionen beinhaltet. Das Verfahren startet dabei mit einer *beliebigen* Teilmenge $\mathcal{J}_0 \subseteq J(x^0)$ und beruht darauf, im $(K + 1)$ -ten Iterationsschritt das Programm

$$\begin{aligned} (\overline{\text{QP}}^K) \quad & \min_{x \in \mathbb{R}^n} \quad \frac{1}{2} x^\top Q x + c^\top x \\ & s.t. \quad A_{\mathcal{J}_K} x = \alpha_{\mathcal{J}_K} \\ & \quad \quad Bx = \beta \end{aligned}$$

zu lösen, und zwar unter Verwendung von Satz 4.1.1 mit $\hat{x} = x^K$. Die Approximationen \mathcal{J}_K von $J(x^K)$ werden dabei iterativ wie folgt bestimmt:

Einerseits wird der Index einer (inaktiven) Ungleichheitsrestriktion in \mathcal{J}_K aufgenommen, wenn ein Iterationsschritt stattfindet, der zur „Aktivierung“ dieser Restriktion führt. Ist andererseits auf der aktuellen Menge \mathcal{J}_K der aktiven Restriktionen keine Verbesserung des Zielfunktionswertes mehr realisierbar, so ermittelt der Algorithmus - falls möglich - eine aktive Ungleichheitsrestriktion, deren „Inaktivierung“ einen besseren Zielfunktionswert ermöglichen würde, und entfernt den korrespondierenden Index aus \mathcal{J}_K . Diese Methodik stellt sicher, dass stets $\mathcal{J}_K \subseteq J(x^K)$ erfüllt bleibt und hat zum Ziel, die tatsächliche Menge der aktiven Restriktionen für das globale Optimum zu finden.

Da das Optimierungsproblem $(\overline{\text{QP}}^K)$ die (in x^K) inaktiven Restriktionen (sowie eventuell auch einige aktive Restriktionen) gänzlich ignoriert, ist eine separate Berücksichtigung

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

dieser Ungleichheitsrestriktionen notwendig, um das Erreichen von unzulässigen Punkten zu vermeiden. Die entsprechende Vorgangsweise ist im nachstehenden Pseudoalgorithmus ersichtlich:

Algorithmus 4.1 Die Methode der aktiven Menge

Input:

- Eine Instanz des quadratischen Optimierungsproblems (QP), welche durch $Q \in \mathbb{R}^{n \times n}$ (symmetrisch und positiv semidefinit), $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{p \times n}$, $\alpha \in \mathbb{R}^p$, $B \in \mathbb{R}^{q \times n}$ und $\beta \in \mathbb{R}^q$ beschrieben wird.
- Ein für (QP) zulässiger Punkt $x^0 \in \mathbb{R}^n$.

Output:

- Ein KKT-Punkt (x^*, λ, γ) von (QP) (womit x^* ein globales Minimum ist).
-

- 1: Setze $\mathcal{J}_0 \subseteq \{i \in \{1, \dots, p\} \mid a_i^\top x^0 = \alpha_i\}$.
- 2: Setze $K := 0$.
- 3: Setze $\lambda_j^{K+1} := 0$ für alle $j \in \mathcal{J}_K^c$ (mit $\mathcal{J}_K^c := \{1, \dots, p\} \setminus \mathcal{J}_K$).
- 4: Berechne $(\Delta x^K, \lambda_{\mathcal{J}_K}^{K+1}, \gamma^{K+1}) \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{J}_K|} \times \mathbb{R}^q$ als Lösung von

$$\begin{pmatrix} Q & -A_{\mathcal{J}_K}^\top & -B^\top \\ A_{\mathcal{J}_K} & 0 & 0 \\ B & 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta x^K \\ \lambda_{\mathcal{J}_K}^{K+1} \\ \gamma^{K+1} \end{pmatrix} = \begin{pmatrix} -Qx^K - c \\ 0 \\ 0 \end{pmatrix}.$$

- 5: **if** $\Delta x^K = 0$ **then**
- 6: **if** $\lambda_j^{K+1} \geq 0$ für alle $j \in \mathcal{J}_K$ **then**
- 7: Setze $x^{K+1} := x^K$.
- 8: **return** $(x^{K+1}, \lambda^{K+1}, \gamma^{K+1})$
- 9: **else if** $\min\{\lambda_j^{K+1} \mid j \in \mathcal{J}_K\} < 0$ **then**
- 10: Berechne einen beliebigen Index r mit $r \in \operatorname{argmin}_{j \in \mathcal{J}_K} \{\lambda_j^{K+1}\}$.
- 11: Setze $x^{K+1} := x^K$.
- 12: Setze $\mathcal{J}_{K+1} := \mathcal{J}_K \setminus \{r\}$.
- 13: **end if**
- 14: **else if** $\Delta x^K \neq 0$ **then**
- 15: **if** $x^K + \Delta x^K$ zulässig für (QP) **then**
- 16: Setze $x^{K+1} := x^K + \Delta x^K$.
- 17: Setze $\mathcal{J}_{K+1} := \mathcal{J}_K$.
- 18: **else if** $x^K + \Delta x^K$ nicht zulässig für (QP) **then**
- 19: Berechne einen beliebigen Index s mit

$$s \in \operatorname{argmin}_{j \in \mathcal{J}_K^c} \left\{ \frac{\alpha_j - a_j^\top x^K}{a_j^\top \Delta x^K} \mid a_j^\top \Delta x^K < 0 \right\}.$$

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

```

20:     Setze  $t^K := \frac{\alpha_s - a_s^\top x^K}{a_s^\top \Delta x^K}$ .
21:     Setze  $x^{K+1} := x^K + t^K \Delta x^K$ .
22:     Setze  $\mathcal{J}_{K+1} := \mathcal{J}_K \cup \{s\}$ .
23:     end if
24: end if
25: Setze  $K := K + 1$ .
26: goto Zeile 3.

```

Wir analysieren nun die Funktionsweise des obigen Algorithmus (siehe Geiger und Kanzow, 2002):

Der Algorithmus startet mit einem für (QP) zulässigen Punkt x^0 . Aus den vier Fallunterscheidungen in den Zeilen 5 bis 24 folgt induktiv die Zulässigkeit des Punktes x^K in der $(K + 1)$ -ten Iteration, weswegen $(x^K + \Delta x^K, \lambda_{\mathcal{J}_K}^{K+1}, \gamma^{K+1})$ gemäß Satz 4.1.1 ein KKT-Punkt von $(\overline{\text{QP}}^K)$ ist:

- *1. Fall* (Zeilen 6 bis 8):
Ist $\Delta x^K = 0$ und $\lambda_j^{K+1} \geq 0, \forall j \in \mathcal{J}_K$, so setzen wir $x^{K+1} := x^K$. Damit haben wir mit $(x^{K+1}, \lambda^{K+1}, \gamma^{K+1})$ (wobei $\lambda_j^{K+1} = 0, \forall j \in \mathcal{J}_K^c$) einen KKT-Punkt von (QP) gefunden und der Algorithmus bricht ab.
- *2. Fall* (Zeilen 9 bis 12):
Unter der Bedingung, dass $\Delta x^K = 0$ ist, jedoch ein Index $j \in \mathcal{J}_K$ mit $\lambda_j^{K+1} < 0$ existiert, ist mit $x^{K+1} := x^K$ der Punkt $(x^{K+1}, \lambda^{K+1}, \gamma^{K+1})$ kein KKT-Punkt von (QP), da hierfür aufgrund der komplementären Schlupfbedingung sämtliche Lagrange-Multiplikatoren, die einer Ungleichheitsrestriktion zuzuordnen sind, nichtnegativ sein müssten. $\Delta x^K = 0$ zeigt uns, dass mit \mathcal{J}_K als Indexmenge der aktiven Ungleichheitsrestriktionen keine Verringerung des Zielfunktionswertes mehr möglich ist. Eine solche Verbesserung ist nur dann möglich, wenn ein Index aus dieser Menge entfernt wird. Für diesen *Inaktivierungsschritt* wählen wir jenen Index r , dessen zugehöriger Lagrange-Multiplikator λ_r^{K+1} am stärksten negativ ist.⁹
- *3. Fall* (Zeilen 15 bis 17):
Falls $\Delta x^K \neq 0$ und $x^K + \Delta x^K$ für (QP) zulässig ist, so setzen wir $x^{K+1} := x^K + \Delta x^K$. Dies ist sinnvoll, da - wie wir später sehen werden - unter der Voraussetzung $Q \succ 0$ der Zielfunktionswert von (QP) im neuen Punkt x^{K+1} strikt kleiner ist als in x^K .
- *4. Fall* (Zeilen 18 bis 22):
Ist $x^K + \Delta x^K$ für (QP) nicht zulässig, bedeutet dies, dass in diesem Punkt zumindest eine der Ungleichheitsrestriktionen, die *nicht* in \mathcal{J}_K enthalten sind, verletzt wird. Das Ziel des Algorithmus ist nun, einen Iterationsschritt $x^{K+1} := x^K + t^K \Delta x^K$ mit

⁹Nocedal und Wright (2006) weisen darauf hin, dass für diese Wahl die größte Verringerungsrate in der Zielfunktion gegeben ist, daraus jedoch nicht zwangsweise die größte absolute Verbesserung des Zielfunktionswertes resultieren muss. Die verbleibenden Restriktionen können nämlich zu einem sehr kleinen Schritt in der nächsten Iteration führen. Die Verwendung des am stärksten negativen Lagrange-Multiplikators ist also heuristisch motiviert. Im Allgemeinen eignet sich jeder Index j mit $\lambda_j^{K+1} < 0$ zur Entfernung aus \mathcal{J}_K .

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

größtmöglichem $t^K \in [0, 1)$ durchzuführen, sodass x^{K+1} zulässig bleibt. Dies bedeutet also, dass

$$a_j^\top x^{K+1} = a_j^\top x^K + t^K a_j^\top \Delta x^K \geq \alpha_j, \quad \forall j \in \mathcal{J}_K^c \quad (4.4)$$

erfüllt sein muss. Wir wissen aus der Zulässigkeit von x^K , dass $a_j^\top x^K \geq \alpha_j, \forall j \in \mathcal{J}_K^c$. Ist nun für einen Index j noch $a_j^\top \Delta x^K \geq 0$, so gilt (4.4) auf jeden Fall. Für $a_j^\top \Delta x^K < 0$ entspricht (4.4) hingegen der Bedingung

$$t^K a_j^\top \Delta x^K \geq \alpha_j - a_j^\top x^K \quad \Leftrightarrow \quad t^K \leq \frac{\alpha_j - a_j^\top x^K}{a_j^\top \Delta x^K}.$$

Daraus folgt die Berechnung von t^K in der 19. und 20. Zeile des Algorithmus. Aus dieser Vorgangsweise ist unmittelbar zu erkennen, dass zumindest eine (zuvor inaktive) Ungleichheitsrestriktion aktiv wird. Wir nehmen daher den Index *einer* solchen Restriktion in die neue Menge \mathcal{J}_{K+1} auf.

Es ist zu beachten, dass \mathcal{J}_{K+1} somit (und wegen Zeile 1 des Algorithmus) auch eine *echte* Teilmenge der Indexmenge der in x^{K+1} aktiven Ungleichheitsrestriktionen sein kann. Aus diesem Grund ist es auch nicht ausgeschlossen, dass die Schrittweite t^K nach obiger Berechnungsmethode den Wert 0 annimmt, worauf wir später noch eingehen werden.

Allerdings ist stets gewährleistet, dass in (4.4) ein Index $j \in \mathcal{J}_K^c$ existiert, sodass $a_j^\top \Delta x^K < 0$ ist. Wäre dies nämlich nicht der Fall, so würde

$$a_j^\top (x^K + \Delta x^K) \geq \alpha_j, \quad \forall j \in \mathcal{J}_K^c$$

folgen. Da sämtliche Gleichheitsrestriktionen sowie jene Ungleichheitsrestriktionen, deren Indizes in \mathcal{J}_K enthalten sind, im Punkt $x^K + \Delta x^K$ wegen des Gleichungssystems in der 4. Zeile und Satz 4.1.1 ohnehin erfüllt sind, wäre $x^K + \Delta x^K$ ein zulässiger Punkt für (QP) und der vorherige 3. Fall würde eintreffen.

Der folgende Satz garantiert nun unter gewissen Voraussetzungen vorteilhafte Eigenschaften des Algorithmus:

Satz 4.1.2. *Es sei ein quadratisches Optimierungsproblem (QP) mit positiv definiten Matrix Q und Zielfunktion $f(x) := \frac{1}{2}x^\top Qx + c^\top x$ gegeben. Dann gelten für Algorithmus 4.1 folgende Aussagen:*

- a) *Das lineare Gleichungssystem in Zeile 4 ist eindeutig lösbar, falls die Vektoren $a_j, \forall j \in \mathcal{J}_K$ und $b_k, \forall k \in \{1, \dots, q\}$ linear unabhängig sind.*
- b) *Falls zu Beginn der $(K+1)$ -ten Iteration die Vektoren $a_j, \forall j \in \mathcal{J}_K$ und $b_k, \forall k \in \{1, \dots, q\}$ linear unabhängig sind und der Algorithmus in dieser Iteration nicht terminiert (also Zeile 8 nicht erreicht), dann sind die Vektoren $a_j, \forall j \in \mathcal{J}_{K+1}$ und $b_k, \forall k \in \{1, \dots, q\}$ ebenfalls linear unabhängig.*

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

c) Der in der $(K + 1)$ -ten Iteration in Zeile 4 bestimmte Vektor Δx^K sei ungleich 0. Dann gilt mit $x^{K+1} := x^K + t^K \Delta x^K$ für alle $t^K \in (0, 1]$, dass

$$f(x^{K+1}) < f(x^K),$$

und Δx^K ist eine Abstiegsrichtung von f in x^K .¹⁰

d) Falls in der $(K + 1)$ -ten Iteration die Vektoren $a_j, \forall j \in \mathcal{J}_K$ und $b_k, \forall k \in \{1, \dots, q\}$ linear unabhängig sind und $\Delta x^K = 0$ sowie $\min\{\lambda_j^{K+1} \mid j \in \mathcal{J}_K\} < 0$ gelten, so ist in der darauffolgenden Iteration $\Delta x^{K+1} \neq 0$.

Beweis.

(Vgl. Geiger und Kanzow (2002) für a), b) und c) sowie Nocedal und Wright (2006) für d.)

a) Wir betrachten folgendes homogene Gleichungssystem:

$$\begin{pmatrix} Q & -A_{\mathcal{J}_K}^\top & -B^\top \\ A_{\mathcal{J}_K} & 0 & 0 \\ B & 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta \hat{x} \\ \hat{\lambda}_{\mathcal{J}_K} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad (4.5)$$

Wir zeigen nun, dass die Lösung $(\Delta \hat{x}, \hat{\lambda}_{\mathcal{J}_K}, \hat{\gamma}) = (0, 0, 0)$ eindeutig ist, woraus unmittelbar das zu Beweisende folgt:

Aus (4.5) ist ersichtlich, dass jede Lösung

$$Q\Delta \hat{x} - A_{\mathcal{J}_K}^\top \hat{\lambda}_{\mathcal{J}_K} - B^\top \hat{\gamma} = 0 \quad (4.6)$$

sowie

$$A_{\mathcal{J}_K} \Delta \hat{x} = 0 \quad \text{und} \quad B \Delta \hat{x} = 0 \quad (4.7)$$

erfüllen muss. Multiplizieren wir nun die erste Gleichung von links mit $\Delta \hat{x}^\top$ und setzen die beiden Bedingungen aus (4.7) ein, so erhalten wir

$$\Delta \hat{x}^\top Q \Delta \hat{x} - (A_{\mathcal{J}_K} \Delta \hat{x})^\top \hat{\lambda}_{\mathcal{J}_K} - (B \Delta \hat{x})^\top \hat{\gamma} = \Delta \hat{x}^\top Q \Delta \hat{x} = 0.$$

Dies ist aufgrund der angenommenen positiven Definitheit von Q nur für $\Delta \hat{x} = 0$ erfüllt. Damit entspricht die Gleichung (4.6) der Bedingung

$$(-A_{\mathcal{J}_K}^\top \quad -B^\top) \begin{pmatrix} \hat{\lambda}_{\mathcal{J}_K} \\ \hat{\gamma} \end{pmatrix} = 0. \quad (4.8)$$

Da laut Annahme die Vektoren $a_j, \forall j \in \mathcal{J}_K$ und $b_k, \forall k \in \{1, \dots, q\}$, welche den (negativen) Spalten der Matrix in (4.8) entsprechen, linear unabhängig sind, müssen sowohl $\hat{\lambda}_{\mathcal{J}_K}$ als auch $\hat{\gamma}$ gleich 0 sein. Insgesamt löst also nur der Vektor $(\Delta \hat{x}, \hat{\lambda}_{\mathcal{J}_K}, \hat{\gamma}) = (0, 0, 0)$ das Gleichungssystem (4.5).

¹⁰Die Definition einer Abstiegsrichtung ist in Appendix A.5 nachzulesen.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

- b) Falls in der $(K + 1)$ -ten Iteration der *2. Fall* (Zeilen 9 bis 12) oder der *3. Fall* (Zeilen 15 bis 17) zur Anwendung kommt, so gilt $\mathcal{J}_{K+1} \subseteq \mathcal{J}_K$. Dies impliziert klarerweise die Aussage des Satzes.

Falls der *4. Fall* (Zeilen 18 bis 22) eintritt, dann ist hingegen

$$|\mathcal{J}_{K+1} \setminus \mathcal{J}_K| = 1.$$

Bezeichnen wir den in \mathcal{J}_{K+1} neu hinzugekommenen Index (wie in Algorithmus 4.1) mit s , so gilt wegen der 19. und 20. Zeile des Algorithmus

$$a_s^\top \Delta x^K < 0 \quad \text{und} \quad t^K = \frac{\alpha_s - a_s^\top x^K}{a_s^\top \Delta x^K}.$$

Wir zeigen die zu beweisende Aussage nun durch Widerspruch:

Sei a_s linear abhängig von $a_j, \forall j \in \mathcal{J}_K$ und $b_k, \forall k \in \{1, \dots, q\}$. Dann existieren Koeffizienten $\delta_j^a, j \in \mathcal{J}_K$ und $\delta_k^b, k \in \{1, \dots, q\}$, sodass gilt:

$$a_s = \sum_{j \in \mathcal{J}_K} \delta_j^a a_j + \sum_{k=1}^q \delta_k^b b_k \quad \Rightarrow \quad a_s^\top \Delta x^K = \sum_{j \in \mathcal{J}_K} \delta_j^a a_j^\top \Delta x^K + \sum_{k=1}^q \delta_k^b b_k^\top \Delta x^K$$

Wegen $A_{\mathcal{J}_K} \Delta x^K = 0$ und $B \Delta x^K = 0$ folgt $a_s^\top \Delta x^K = 0$. Dies ist jedoch ein Widerspruch zu $a_s^\top \Delta x^K < 0$, womit die Annahme der linearen Abhängigkeit fallengelassen werden muss.

- c) Wir untersuchen analog zum homogenen Gleichungssystem aus Teil a) des Beweises das inhomogene System aus Zeile 4 des Algorithmus und erhalten

$$Q \Delta x^K - A_{\mathcal{J}_K}^\top \lambda_{\mathcal{J}_K}^{K+1} - B^\top \gamma^{K+1} = -Q x^K - c$$

\Rightarrow

$$(\Delta x^K)^\top Q \Delta x^K - (A_{\mathcal{J}_K} \Delta x^K)^\top \lambda_{\mathcal{J}_K}^{K+1} - (B \Delta x^K)^\top \gamma^{K+1} = -(\Delta x^K)^\top Q x^K - c^\top \Delta x^K$$

sowie

$$A_{\mathcal{J}_K} \Delta x^K = 0 \quad \text{und} \quad B^\top \Delta x^K = 0,$$

woraus

$$(\Delta x^K)^\top Q \Delta x^K = -(\Delta x^K)^\top Q x^K - c^\top \Delta x^K \quad (4.9)$$

resultiert.¹¹ Für $t^K = 1$ und $\Delta x^K \neq 0$ folgt daher

$$f(x^K + \Delta x^K) - f(x^K) = \frac{1}{2}(x^K)^\top Q x^K + 2 \cdot \frac{1}{2}(\Delta x^K)^\top Q x^K + \frac{1}{2}(\Delta x^K)^\top Q \Delta x^K$$

¹¹Laut Annahme ist Q positiv definit und $\Delta x^K \neq 0$, womit wegen

$$\nabla f(x^K)^\top \Delta x^K = (\Delta x^K)^\top Q x^K + c^\top \Delta x^K \stackrel{(4.9)}{=} -(\Delta x^K)^\top Q \Delta x^K < 0$$

hier bereits klar ist, dass Δx^K eine Abstiegsrichtung von f in x^K ist (Lemma A.5.1).

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

$$\begin{aligned}
& + c^\top x^K + c^\top \Delta x^K - \frac{1}{2}(x^K)^\top Q x^K - c^\top x^K \\
& = (\Delta x^K)^\top Q x^K + \frac{1}{2}(\Delta x^K)^\top Q \Delta x^K + c^\top \Delta x^K \\
& \stackrel{(4.9)}{=} -\frac{1}{2}(\Delta x^K)^\top Q \Delta x^K \\
& < 0,
\end{aligned}$$

also

$$f(x^K + \Delta x^K) < f(x^K). \quad (4.10)$$

Da die Zielfunktion f aufgrund der positiven Definitheit von Q strikt konvex ist, lässt sich nun auch für $t^K \in (0, 1)$ die Aussage des Satzes folgern. Es gilt nämlich

$$\begin{aligned}
f(x^K + t^K \Delta x^K) & = f((1 - t^K)x^K + t^K(x^K + \Delta x^K)) \\
& < (1 - t^K)f(x^K) + t^K f(x^K + \Delta x^K) \\
& \stackrel{(4.10)}{<} (1 - t^K)f(x^K) + t^K f(x^K) \\
& = f(x^K).
\end{aligned}$$

Damit ist Δx^K per Definition eine Abstiegsrichtung (vgl. Appendix A.5).

- d) Nach den Voraussetzungen des Satzes befinden wir uns im 2. Fall (Zeilen 9 bis 12) des Algorithmus. Sei r jener Index, der in der aktuellen $(K + 1)$ -Iteration aus \mathcal{J}_K entfernt wird (d.h. $\lambda_r^{K+1} < 0$ und $\mathcal{J}_{K+1} = \mathcal{J}_K \setminus \{r\}$). Betrachten wir das Gleichungssystem in Zeile 4, so gilt wegen $\Delta x^K = 0$ (und damit $x^{K+1} = x^K$)

$$-A_{\mathcal{J}_K}^\top \lambda_{\mathcal{J}_K}^{K+1} - B^\top \gamma^{K+1} = -Qx^K - c$$

sowie

$$Q\Delta x^{K+1} - A_{\mathcal{J}_{K+1}}^\top \lambda_{\mathcal{J}_{K+1}}^{K+2} - B^\top \gamma^{K+2} = -Qx^K - c.$$

Subtrahieren wir diese beiden Gleichungen voneinander, so folgt

$$Q\Delta x^{K+1} - \sum_{j \in \mathcal{J}_{K+1}} a_j (\lambda_j^{K+2} - \lambda_j^{K+1}) + a_r \lambda_r^{K+1} - B^\top (\gamma^{K+2} - \gamma^{K+1}) = 0.$$

Nehmen wir nun an, dass $\Delta x^{K+1} = 0$ ist, so gilt

$$a_r \lambda_r^{K+1} = \sum_{j \in \mathcal{J}_{K+1}} a_j (\lambda_j^{K+2} - \lambda_j^{K+1}) + B^\top (\gamma^{K+2} - \gamma^{K+1}).$$

Dies ist wegen $\lambda_r^{K+1} < 0$ jedoch ein Widerspruch zur linearen Unabhängigkeit von $a_j, \forall j \in \mathcal{J}_K$ und $b_k, \forall k \in \{1, \dots, q\}$.

□

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

Damit ist offenbar das Gleichungssystem in Zeile 4 des Algorithmus der aktiven Menge stets eindeutig lösbar, wenn folgende Voraussetzungen erfüllt sind:

- i) Die Matrix Q ist positiv definit.
- ii) Die Zeilen der Matrix B , welche die Gleichheitsrestriktionen beschreiben, sind linear unabhängig, was gegebenenfalls durch Eliminierung überflüssiger Restriktionen erreicht werden kann.
- iii) Die Indexmenge \mathcal{J}_0 in Zeile 1 des Algorithmus wird so gewählt, dass $a_j, \forall j \in \mathcal{J}_0$ und $b_k, \forall k \in \{1, \dots, q\}$ linear unabhängig sind.¹²

Wir können nun gemäß Nocedal und Wright (2006) den nachstehenden Satz über die Endlichkeit des Verfahrens der aktiven Menge formulieren:

Satz 4.1.3. *Es sei ein quadratisches Optimierungsproblem (QP) für die Methode der aktiven Menge gegeben. Weiters seien die obigen drei Bedingungen i), ii) und iii) sowie die Annahme, dass die Schrittweite t^K in jeder Iteration strikt positiv ist, erfüllt. Dann gelingt es Algorithmus 4.1 in endlich vielen Iterationen eine Optimallösung zu bestimmen.*

Beweis. (Siehe Nocedal und Wright, 2006.)

Wir zeigen die folgenden beiden Aussagen:

Behauptung: Spätestens in jeder $(n - q)$ -ten Iteration tritt der Fall $\Delta x^K = 0$ ein.

Beweis: Ist mit $\Delta x^K \neq 0$ der Punkt $x^K + \Delta x^K$ für (QP) zulässig, wählt der Algorithmus $x^{K+1} := x^K + \Delta x^K$ und Δx^{K+1} ist im nächsten Iterationsschritt wegen Satz 4.1.1 gleich 0. Falls $x^K + \Delta x^K$ hingegen unzulässig ist, wird eine kürzere Schrittweite $t^K < 1$ berechnet und ein zusätzlicher Index zur Menge \mathcal{J}_K hinzugenommen. Tritt dieser Fall wiederholt ein, so ist jedoch nach spätestens $l \leq n - q$ Iterationen $|\mathcal{J}_{K+l}| + q = n$. Die Matrix $\begin{pmatrix} A_{\mathcal{J}_{K+l}} \\ B \end{pmatrix}$ ist damit wegen der linearen Unabhängigkeit der Zeilenvektoren eine reguläre $(n \times n)$ -Matrix, weshalb für die einzige Lösung des Gleichungssystems in Zeile 4 des Algorithmus $\Delta x^{K+l} = 0$ gelten muss.

Behauptung: Falls $\Delta x^K = 0$ ist und ein Element von \mathcal{J}_K gelöscht wird (2. Fall), so erreicht der Algorithmus die Menge \mathcal{J}_K in späteren Iterationen nicht mehr.

Beweis: Ist in der $(K + 1)$ -ten Iteration des Algorithmus $\Delta x^K = 0$, so ist wegen Satz 4.1.1 der Punkt x^K klarerweise das strikt globale Minimum von $(\overline{\text{QP}}^K)$. Falls der Algorithmus nicht terminiert und damit x^K noch keine Optimallösung von (QP) ist, so ist aufgrund von Satz 4.1.2, Teil d) der in der darauffolgenden Iteration berechnete Wert $\Delta x^{K+1} \neq 0$. Gemeinsam mit der Annahme $t^K > 0$ folgt daher aus Teil c) desselben Satzes, dass $f(x_{K+l}) < f(x_{K+1}) = f(x_K), \forall l > 1$. Deswegen und aufgrund der Tatsache, dass x^K ein globales Optimum des Problems $(\overline{\text{QP}}^K)$ ist, sind alle Punkte x_{K+l} mit $l > 1$ für $(\overline{\text{QP}}^K)$

¹²Dies bedeutet also, dass es unter Umständen auch wirklich sinnvoll sein kann, \mathcal{J}_0 als echte Teilmenge der Indizes der in x^0 aktiven Ungleichheitsrestriktionen zu wählen.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

unzulässig. Der Algorithmus kann also im weiteren Verlauf nicht mehr zur Menge \mathcal{J}_K zurückkehren.

Aufgrund dieser Beobachtungen und der Tatsache, dass die Anzahl p der Ungleichheitsrestriktionen und damit auch die maximale Anzahl der möglichen Mengen \mathcal{J}_K endlich ist, muss der Algorithmus offensichtlich in einer endlichen Anzahl von Iterationen bei der Optimallösung ankommen. \square

Dass die Annahme $t^K > 0$ im obigen Satz jedoch nicht zwangsläufig erfüllt sein muss, ergibt sich unmittelbar aus der Tatsache, dass die Komplementärmenge \mathcal{J}_K^c auch Indizes von (in x^K) aktiven Ungleichheitsrestriktionen beinhalten kann. Es können also „Zyklen“ zwischen dem 2. Fall (Zeilen 9 bis 12) und dem 4. Fall (Zeilen 18 bis 22) mit $x^{K+l} = x^K$ und $\mathcal{J}_{K+l} = \mathcal{J}_K$ (also ohne Zielfunktionswertveränderung) auftreten. Nocedal und Wright (2006) verweisen darauf, dass es möglich ist, dieses Problem analog zu den Methoden für das Simplexverfahren in der linearen Programmierung zu behandeln. Geiger und Kanzow (2002) merken an, dass in der Praxis von einer endlichen Iterationszahl ausgegangen werden kann.

4.1.2.1. Die Berechnung einer zulässigen Ausgangslösung

Wir besprechen nun noch eine Möglichkeit zur Bestimmung eines zulässigen Punktes für ein quadratisches Optimierungsproblem, da ein solcher für das Verfahren der aktiven Menge benötigt wird. Die folgende Vorgangsweise wird z.B. in Nocedal und Wright (2006) mit einem ähnlichen Hilfsproblem vorgeschlagen. Die verwendeten Begriffe und Sätze aus der linearen Optimierung sind beispielsweise in Geiger und Kanzow (2002) nachzulesen:

Lemma 4.1.4. *Wir betrachten ein quadratisches Optimierungsproblem (QP) mit den Restriktionen $Ax \geq \alpha$ und $Bx = \beta$ sowie das lineare Optimierungsproblem*

$$\begin{aligned}
 \min_{\substack{z^\alpha \in \mathbb{R}^p \\ z^\beta \in \mathbb{R}^q}} & e^\top z^\alpha + e^\top z^\beta \\
 \text{s.t.} & a_i^\top x^+ - a_i^\top x^- - y_i + \delta_i^\alpha z_i^\alpha = \alpha, & i = 1, \dots, p \\
 & b_k^\top x^+ - b_k^\top x^- + \delta_k^\beta z_k^\beta = \beta, & k = 1, \dots, q \\
 & x^+ \geq 0 \\
 & x^- \geq 0 \\
 & y \geq 0 \\
 & z^\alpha \geq 0 \\
 & z^\beta \geq 0
 \end{aligned} \tag{4.11}$$

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

mit $\delta_i^\alpha := \text{sign}(\alpha_i), \forall i = 1, \dots, p$ und $\delta_k^\beta := \text{sign}(\beta_k), \forall k = 1, \dots, q$.¹³

Sei $\begin{pmatrix} \hat{x}^+ \\ \hat{y} \\ \hat{z}^\alpha \\ \hat{z}^\beta \end{pmatrix} \in \mathbb{R}^{n+n+p+p+q}$ ein optimaler Basisvektor von (4.11). Dann gilt:

- a) Ist $\begin{pmatrix} \hat{z}^\alpha \\ \hat{z}^\beta \end{pmatrix} = 0$, so ist $\hat{x} := \hat{x}^+ - \hat{x}^-$ zulässig für (QP).
- b) Ist $\begin{pmatrix} \hat{z}^\alpha \\ \hat{z}^\beta \end{pmatrix} \neq 0$, so besitzt (QP) keinen zulässigen Punkt.

Beweis. (Siehe Geiger und Kanzow (2002) für die Beweisidee.)

- a) Wegen $\hat{y} \geq 0$ folgt mit $\hat{x} = \hat{x}^+ - \hat{x}^-$ aus den Restriktionen von (4.11) unmittelbar $A\hat{x} \geq \alpha$ und $B\hat{x} = \beta$, womit \hat{x} zulässig für (QP) ist.

- b) Wir nehmen an, dass (QP) einen zulässigen Punkt \bar{x} besitzt. Dann ist $\begin{pmatrix} \bar{x}^+ \\ \bar{y} \\ \bar{z}^\alpha \\ \bar{z}^\beta \end{pmatrix}$ mit

$\bar{x}_i^+ := \max(\bar{x}_i, 0) \geq 0, \bar{x}_i^- := -\min(\bar{x}_i, 0) \geq 0, \bar{y}_i := a_i^\top \bar{x} - \alpha_i \geq 0, \bar{z}^\alpha := 0$ und $\bar{z}^\beta := 0$ ein zulässiger Punkt für (4.11) mit Zielfunktionswert $e^\top \bar{z}^\alpha + e^\top \bar{z}^\beta = 0$. Dies ist

jedoch ein Widerspruch zur Annahme, dass $\begin{pmatrix} \hat{x}^+ \\ \hat{x}^- \\ \hat{y} \\ \hat{z}^\alpha \\ \hat{z}^\beta \end{pmatrix}$ optimal ist, da $e^\top \hat{z}^\alpha + e^\top \hat{z}^\beta > 0$.

□

Das lineare Optimierungsproblem (4.11) liegt in der (von Geiger und Kanzow (2002) verwendeten) kanonischen Form $\min \{ \tilde{c}^\top \tilde{x} \mid \tilde{A}\tilde{x} = \tilde{b}, \tilde{x} \geq 0 \}$ vor, wobei $\tilde{A} = \begin{pmatrix} A & -A & -I_p & D_p^\alpha & 0 \\ B & -B & 0 & 0 & D_q^\beta \end{pmatrix}$ mit $D_p^\alpha = \text{diag}(\delta_1^\alpha, \dots, \delta_p^\alpha)$ und $D_q^\beta = \text{diag}(\delta_1^\beta, \dots, \delta_q^\beta)$. Da die Matrix \tilde{A} offensichtlich vollen Zeilenrang hat (d.h. $\text{rank}(\tilde{A}) = p + q$) und $(0^\top, 0^\top, 0^\top, |\alpha|^\top, |\beta|^\top)^\top$ ein trivialer (Start-)Basisvektor von (4.11) ist, lässt sich dieses lineare Programm problemlos mit dem Simplexverfahren lösen.¹⁴

4.1.3. Innere-Punkte-Verfahren für konvexe Probleme mit Ungleichheitsrestriktionen

Wir möchten nun noch eine zweite, sehr effiziente Methode zur Lösung konvexer quadratischer Probleme untersuchen: die Innere-Punkte-Verfahren. Der Name dieser iterativen Ansätze geht dabei auf die Tatsache zurück, dass sämtliche Ungleichheitsrestriktionen in jeder Iteration *strikt* erfüllt sein müssen. Die folgenden Ausführungen basieren (sofern nicht anders angegeben) auf Nocedal und Wright (2006) sowie Cornuéjols und Tütüncü (2007).

¹³ $\text{sign}(x) := \begin{cases} +1, & \text{falls } x \geq 0 \\ -1, & \text{falls } x < 0 \end{cases}$

¹⁴Die hier beschriebene Vorgangsweise entspricht einer Methode zum Auffinden einer zulässigen Basislösung für beliebige lineare Optimierungsprobleme, welche in der Literatur oftmals als „1. Phase“ des Simplexverfahrens bezeichnet wird.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

Wir betrachten zu Beginn das primale Problem $(\widetilde{\text{QP}})$ sowie dessen Wolfe-duales Problem, welches in Abschnitt 1.5 hergeleitet wurde:

$$\begin{array}{ll}
 (\widetilde{\text{QP}}) & \min_{x \in \mathbb{R}^n} \quad \frac{1}{2} x^\top Q x + c^\top x \\
 & \text{s.t.} \quad Bx = \beta \\
 & \quad \quad x \geq 0 \\
 (\widetilde{\text{QP}})^{\text{W-d}} & \max_{\substack{x \in \mathbb{R}^n \\ \gamma \in \mathbb{R}^q}} \quad -\frac{1}{2} x^\top Q x + \beta^\top \gamma \\
 & \text{s.t.} \quad Qx + c - \lambda - B^\top \gamma = 0 \\
 & \quad \quad \lambda \geq 0
 \end{array}$$

Dabei sind $Q \in \mathbb{R}^{n \times n}$ (o.B.d.A. symmetrisch), $c \in \mathbb{R}^n$, $B \in \mathbb{R}^{q \times n}$ und $\beta \in \mathbb{R}^q$. Wie wir schon in Abschnitt 1.3 gesehen haben, lässt sich jedes Optimierungsproblem der Form (QP) in $(\widetilde{\text{QP}})$ transformieren und kann daher ebenfalls als Input für die folgenden Verfahren dienen.

Wir nehmen wieder (wie bereits für die Methode der aktiven Menge) an, dass $Q \succcurlyeq 0$. Die (hinreichenden) KKT-Bedingungen von $(\widetilde{\text{QP}})$ ergeben sich aus den KKT-Bedingungen [QP-i], [QP-ii] und [QP-iii] von (QP) mit $A = I_n$ und $\alpha = 0$:

- a) $Bx^* - \beta = 0$,
- b) $Qx^* + c - \lambda - B^\top \gamma = 0$,
- c) $x_i \lambda_i = 0, \forall i = 1, \dots, n$,
- d) $x^* \geq 0$,
- e) $\lambda \geq 0$.

Diese Sammlung von Bedingungen lässt sich gemäß Cornuéjols und Tütüncü (2007) auch folgendermaßen interpretiert:

- a) und d) entsprechen der primalen Zulässigkeit
- b) und e) beschreiben die (Wolfe-)duale Zulässigkeit
- c) ist die komplementäre Schlupfbedingung

Die folgenden iterativen Innere-Punkte-Verfahren zielen nun darauf ab, eine Lösung dieses Gleichungssystems zu finden und werden daher auch primal-duale Innere-Punkte-Verfahren genannt. Definieren wir die Funktion

$$\begin{aligned}
 F : \mathbb{R}^{2n+q} &\rightarrow \mathbb{R}^{2n+q} \\
 (x, \lambda, \gamma) &\mapsto \begin{pmatrix} Qx + c - \lambda - B^\top \gamma \\ Bx - \beta \\ X\Lambda e \end{pmatrix}
 \end{aligned}$$

mit

$$\begin{aligned}
 X &:= \text{diag}(x_1, \dots, x_n), \\
 \Lambda &:= \text{diag}(\lambda_1, \dots, \lambda_n), \\
 e &:= (1, \dots, 1)^\top \in \mathbb{R}^n,
 \end{aligned}$$

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

so entsprechen die Optimalitätskriterien a) bis c) der Forderung

$$F(x^*, \lambda, \gamma) = 0 \in \mathbb{R}^{2n+q}.$$

Für deren Lösung wird im Wesentlichen auf das in Appendix A.6 beschriebene Newton-Verfahren zurückgegriffen. Liegt also ein aktueller Iterationspunkt x^K vor, so wird eine Taylorapproximation 1. Ordnung, nämlich

$$F(x^K + \Delta x^K, \lambda^K + \Delta \lambda^K, \gamma^K + \Delta \gamma^K) \approx F(x^K, \lambda^K, \gamma^K) + \mathcal{J}_F(x^K, \lambda^K, \gamma^K) \cdot \begin{pmatrix} \Delta x^K \\ \Delta \lambda^K \\ \Delta \gamma^K \end{pmatrix}$$

durchgeführt. Nach Berechnung der Jacobi-Matrix und unter Verwendung von

$$\begin{aligned} X^K &:= \text{diag}(x_1^K, \dots, x_n^K), \\ \Lambda^K &:= \text{diag}(\lambda_1^K, \dots, \lambda_n^K), \end{aligned}$$

(analog zu den Diagonalmatrizen X und Λ) sowie

$$F(x^K, \lambda^K, \gamma^K) = \begin{pmatrix} Qx^K + c - \lambda^K - B^\top \gamma^K \\ Bx^K - \beta \\ X^K \Lambda^K e \end{pmatrix} =: \begin{pmatrix} r_d^K \\ r_p^K \\ X^K \Lambda^K e \end{pmatrix}$$

für das primale und das duale Residuum erhalten wir durch Nullsetzen der obigen Approximation das folgende lineare Gleichungssystem:

$$\begin{pmatrix} Q & -I_n & -B^\top \\ B & 0 & 0 \\ \Lambda^K & X^K & 0 \end{pmatrix} \cdot \begin{pmatrix} \Delta x^K \\ \Delta \lambda^K \\ \Delta \gamma^K \end{pmatrix} = \begin{pmatrix} -r_d^K \\ -r_p^K \\ -X^K \Lambda^K e \end{pmatrix}$$

Allerdings ist zu beachten, dass in diesem Schema die beiden Nichtnegativitätsbedingungen d) und e) noch nicht berücksichtigt wurden. Daher ist es im Allgemeinen nicht (wie im zugrundeliegenden Newton-Verfahren) möglich,

$$(x^{K+1}, \lambda^{K+1}, \gamma^{K+1}) := (x^K, \lambda^K, \gamma^K) + t^K (\Delta x^K, \Delta \lambda^K, \Delta \gamma^K)$$

mit $t^K = 1$ zu setzen. Stattdessen wird $t^K \in (0, 1]$ so bestimmt, dass $x^{K+1} \geq 0$ und $\lambda^{K+1} \geq 0$ sogar *strikt* erfüllt sind. Cornuéjols und Tütüncü (2007) weisen darauf hin, dass die entsprechenden größtmöglichen Werte analog zum sogenannten „Ratio Test“ des Simplexverfahrens für lineare Optimierungsprobleme ermittelt werden können.

Löst man obiges Gleichungssystem, so erhält man einen Vektor, welcher *reine Newton-Richtung* („pure Newton direction“) genannt wird und in der Praxis oftmals zu sehr kleinen Schrittweiten t^K führt. Um diesen Nachteil zu vermeiden, werden nachstehende Modifikationen vorgenommen, welche zu den *pfadverfolgenden* Verfahren führen:

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

Ziel ist es, die Bedingung c) so abzuschwächen, dass $x_i^K \lambda_i^K$ einen (für alle $i = 1, \dots, n$ konstanten) positiven Wert annimmt. Dafür definieren wir das **Dualitätsmaß** der beiden Vektoren $x^K \in \mathbb{R}^n$ und $\lambda^K \in \mathbb{R}^n$ als

$$\mu^K := \frac{(x^K)^\top \lambda^K}{n}$$

und führen in der $(K+1)$ -ten Iteration wieder eine Taylorapproximation 1. Ordnung um den Punkt $(x^K, \lambda^K, \gamma^K)$ durch, nun allerdings von der modifizierten Funktion

$$\hat{F}(x^K + \Delta x^K, \lambda^K + \Delta \lambda^K, \gamma^K + \Delta \gamma^K) := F(x^K + \Delta x^K, \lambda^K + \Delta \lambda^K, \gamma^K + \Delta \gamma^K) - \begin{pmatrix} 0 \\ \sigma^K \mu^K e \end{pmatrix}$$

mit $\sigma^K \in [0, 1]$. Es ist also das Gleichungssystem

$$\begin{pmatrix} Q & -I_n & -B^\top \\ B & 0 & 0 \\ \Lambda^K & X^K & 0 \end{pmatrix} \cdot \begin{pmatrix} \Delta x^K \\ \Delta \lambda^K \\ \Delta \gamma^K \end{pmatrix} = \begin{pmatrix} -r_d^K \\ -r_p^K \\ -X^K \Lambda^K e + \sigma^K \mu^K e \end{pmatrix} \quad (4.12)$$

zu lösen, wobei σ^K jener gewünschte Faktor ist, um den sich das Dualitätsmaß, welches gewissermaßen den „Abstand“ vom Zustand $x_i \lambda_i = 0, \forall i = 1, \dots, n$ angibt, im aktuellen Iterationsschritt reduzieren soll.¹⁵ In der Praxis zeigt sich, dass mit $\sigma^K > 0$ meist größere Schrittweiten t^K realisierbar sind als mit $\sigma^K = 0$.

Mit diesen Ideen können wir nun folgenden generischen Algorithmus formulieren:

Algorithmus 4.2 *Ein generisches pfadverfolgendes Innere-Punkte-Verfahren*

Input:

- Eine Instanz des quadratischen Optimierungsproblems (\widetilde{QP}) , welche durch $Q \in \mathbb{R}^{n \times n}$ (symmetrisch und positiv semidefinit), $c \in \mathbb{R}^n$, $B \in \mathbb{R}^{q \times n}$ und $\beta \in \mathbb{R}^q$ beschrieben wird.
- Ein Punkt $(x^0, \lambda^0, \gamma^0) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^q$ mit $x^0 > 0$ und $\lambda^0 > 0$.
- Ein Abbruchkriterium.

Output:

- Ein approximativer KKT-Punkt $(\hat{x}, \lambda, \gamma)$ von (\widetilde{QP}) (womit \hat{x} ein approximatives globales Minimum ist).
-

- 1: Setze $K := 0$.
- 2: Setze $r_p^K := Bx^K - \beta$, $r_d^K := Qx^K + c - \lambda^K - B^\top \gamma^K$ und $\mu^K := \frac{1}{n}(x^K)^\top \lambda^K$.
- 3: **if** Abbruchkriterium ist erfüllt. **then**
- 4: **return** $(x^K, \lambda^K, \gamma^K)$

¹⁵Die Reduktion entspricht (selbst bei voller Schrittweite) jedoch im Allgemeinen nicht diesem Faktor, da ja eine Linearisierung durchgeführt wird. Diesen Umstand werden wir später im Rahmen des „Predictor-Corrector“-Verfahrens berücksichtigen.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

5: **end if**

6: Wähle $\sigma^K \in [0, 1]$ und berechne $(\Delta x^K, \Delta \lambda^K, \Delta \gamma^K)$ als Lösung von (4.12).

7: Wähle $t^K \in (0, 1]$, sodass $x^K + t^K \Delta x^K > 0$ und $\lambda^K + t^K \Delta \lambda^K > 0$.

8: Setze

$$(x^{K+1}, \lambda^{K+1}, \gamma^{K+1}) := (x^K, \lambda^K, \gamma^K) + t^K (\Delta x^K, \Delta \lambda^K, \Delta \gamma^K).$$

9: Setze $K := K + 1$.

10: **goto** Zeile 2.

Wir zeigen noch folgende Aussage:

Satz 4.1.5. *Es sei ein quadratisches Optimierungsproblem (\widetilde{QP}) mit positiv semidefiniter Matrix Q gegeben. Falls zusätzlich die Matrix B vollen Zeilenrang hat (d.h. $\text{rank}(B) = q$), so ist das in jeder Iteration von Algorithmus 4.2 auftretende lineare Gleichungssystem (4.12) eindeutig lösbar.*

Beweis. Wir orientieren uns an der Vorgangsweise in Geiger und Kanzow (2002) für lineare Optimierungsprobleme und betrachten das entsprechende homogene Gleichungssystem

$$Q \Delta x^K - I_n \Delta \lambda^K - B^\top \Delta \gamma^K = 0, \quad (4.13)$$

$$B \Delta x^K = 0, \quad (4.14)$$

$$\Lambda^K \Delta x^K + X^K \Delta \lambda^K = 0. \quad (4.15)$$

Multiplizieren wir (4.13) von links mit $(\Delta x^K)^\top$, so erhalten wir

$$(\Delta x^K)^\top Q \Delta x^K - (\Delta x^K)^\top \Delta \lambda^K - (\Delta x^K)^\top B^\top \Delta \gamma^K = 0. \quad (4.16)$$

Wegen (4.14) gilt $(\Delta x^K)^\top B^\top = 0$. Darüber hinaus folgt aus (4.15), dass

$$\Delta \lambda^K = -(X^K)^{-1} \Lambda^K \Delta x^K. \quad (4.17)$$

Setzen wir dies in (4.16) ein, so resultiert wegen $Q \succcurlyeq 0$

$$(\Delta x^K)^\top Q \Delta x^K = -(\Delta x^K)^\top (X^K)^{-1} \Lambda^K \Delta x^K \geq 0.$$

Da jedoch $(X^K)^{-1} \Lambda^K$ wegen $x^K > 0$ sowie $\lambda^K > 0$ eine positiv definite Diagonalmatrix ist, folgt $\Delta x^K = 0$ und wegen (4.17) auch $\Delta \lambda^K = 0$. (4.13) liefert schlussendlich (wegen des vollen Zeilenrangs von B) $\Delta \gamma^K = 0$. \square

Wir möchten nun zwei Varianten dieses generischen Algorithmus etwas genauer besprechen:

- Zulässige pfadverfolgende Innere-Punkte-Verfahren, die interessante Konvergenzeigenschaften besitzt.
- Ein unzulässiges Innere-Punkte-Verfahren, welches sich gut zur praktischen Implementierung eignet.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

4.1.3.1. Zulässige pfadverfolgende Innere-Punkte-Verfahren

Wir benötigen die folgenden beiden Mengen:

$$\begin{aligned}\mathcal{F} &:= \{(x, \lambda, \gamma) \mid Bx = \beta, Qx + c - \lambda - B^\top \gamma = 0, x \geq 0, \text{ und } \lambda \geq 0\} \\ \mathcal{F}^\circ &:= \{(x, \lambda, \gamma) \mid Bx = \beta, Qx + c - \lambda - B^\top \gamma = 0, x > 0, \text{ und } \lambda > 0\}\end{aligned}$$

\mathcal{F} entspricht also der Menge der Punkte, welche sowohl primal als auch (Wolfe-)dual zulässig sind, und \mathcal{F}° denjenigen Punkten aus \mathcal{F} , welche die beiden Nichtnegativitätsbedingungen strikt erfüllen. \mathcal{F} heißt daher *zulässige Menge* und \mathcal{F}° *strikt zulässige Menge*.

Zulässige Verfahren sind nun dadurch bestimmt, dass sie in einem Punkt $(x^0, \lambda^0, \gamma^0) \in \mathcal{F}^\circ$ starten, woraus sich $r_d^0 = 0$ und $r_p^0 = 0$ ergibt. Aus (4.12) ist ersichtlich, dass damit (und mit geeignet gewähltem t^K) auch $(x^{K+1}, \lambda^{K+1}, \gamma^{K+1}) \in \mathcal{F}^\circ$ sowie $r_d^{K+1} = 0$ und $r_p^{K+1} = 0$ für alle $K \geq 0$ gilt. Der Algorithmus bewegt sich also ausschließlich auf der strikt zulässigen Menge \mathcal{F}° und in der $(K + 1)$ -ten Iteration ist das Gleichungssystem

$$\begin{pmatrix} Q & -I_n & -B^\top \\ B & 0 & 0 \\ \Lambda^K & X^K & 0 \end{pmatrix} \cdot \begin{pmatrix} \Delta x^K \\ \Delta \lambda^K \\ \Delta \gamma^K \end{pmatrix} =: \begin{pmatrix} 0 \\ 0 \\ -X^K \Lambda^K e + \sigma^K \mu^K e \end{pmatrix} \quad (4.18)$$

zu lösen.

Die Kombination aus zulässigen und pfadverfolgenden Innere-Punkte-Methoden führt nun zum Begriff des **zentralen Pfades**. Dieser ist definiert als

$$\mathcal{C} := \{(x, \lambda, \gamma) \in \mathcal{F}^\circ \mid X \Lambda e = \tau e \text{ mit } \tau > 0\}.$$

Damit erfüllen also die Punkte am zentralen Pfad sämtliche KKT-Bedingungen mit der Ausnahme, dass $x_i \lambda_i = \tau$, $\forall i = 1, \dots, n$ für ein $\tau > 0$ gelten muss. Die Menge \mathcal{C} kann alternativ auch dadurch definiert werden, dass jeder Punkt $(x, \lambda, \gamma) \in \mathcal{C}$ den Bedingungen

$$F(x, \lambda, \gamma) = (0^\top, 0^\top, \tau e^\top)^\top \text{ sowie } x > 0 \text{ und } \lambda > 0 \quad (4.19)$$

für ein $\tau > 0$ genügt. Unter der Voraussetzung, dass $\mathcal{F}^\circ \neq \emptyset$ ist (und B vollen Zeilenrang hat), existiert für jedes $\tau > 0$ sogar eine *eindeutige* Lösung $(x_\tau, \lambda_\tau, \gamma_\tau)$ von (4.19) (siehe Schade, 2008).¹⁶ Der hier verwendete Parameter τ entspricht genau $\sigma^K \mu^K$ in der $(K + 1)$ -ten Iteration von Algorithmus 4.2. Während wir im Fall $\sigma^K = 0$ von einer *reinen* Newton-Richtung gesprochen haben, nennen wir nun im Fall $\sigma^K = 1$ (also $\tau = \mu^K$) die Lösung des Gleichungssystems *zentrierende Newton-Richtung*. Dieser Vektor zeigt auf den Punkt $(x_{\mu^K}, \lambda_{\mu^K}, \gamma_{\mu^K})$ am zentralen Pfad. Die Verwendung dieser Richtung bewirkt klarerweise keine signifikante Verkleinerung des Dualitätsmaßes. Allerdings führt ein Schritt in Richtung

¹⁶Der Beweis dieser Aussage kann unter Verwendung einer Hilfsfunktion mit logarithmischer Schranke erfolgen.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

des zentralen Pfades im Allgemeinen dazu, dass in der darauffolgenden Iteration deutlich größere Verbesserungen möglich sind (vgl. Nocedal und Wright, 2006).

Der steuernde Parameter σ^K heißt daher auch *zentrierender Parameter*. Zulässige pfadverfolgende Ansätze versuchen, die Vor- und Nachteile der reinen und der zentrierenden Newton-Richtung durch eine geeignete Wahl von σ^K zu kombinieren. Dies geschieht oftmals dadurch, dass man sich bei der Bestimmung der Schrittweite t^K auf eine Umgebung des zentralen Pfades konzentriert und die größtmögliche Schrittweite verwendet, die auf der gewählten Umgebung realisierbar ist. Wir betrachten hierzu die folgenden beiden Mengen, welche als **Nachbarschaften des zentralen Pfades** bezeichnet werden:

$$\mathcal{N}_2(\theta) = \left\{ (x, \lambda, \gamma) \in \mathcal{F}^\circ \mid \|X\Lambda e - \mu e\|_2 \leq \theta\mu, \mu = \frac{1}{n}x^\top \lambda \right\} \text{ mit } \theta \in (0, 1)$$

$$\mathcal{N}_S(\psi) = \left\{ (x, \lambda, \gamma) \in \mathcal{F}^\circ \mid \psi\mu \leq x_i\lambda_i \leq \frac{1}{\psi}\mu, \forall i = 1, \dots, n, \mu = \frac{1}{n}x^\top \lambda \right\} \text{ mit } \psi \in (0, 1)$$

In $\mathcal{N}_2(\theta)$, welche wir auch *euklidische* Nachbarschaft nennen, wird also die Menge der zum zentralen Pfad benachbarten Punkte durch die euklidische Norm festgelegt, während dies in der *symmetrischen* Nachbarschaft $\mathcal{N}_S(\psi)$, welche wir gemäß Gondzio (2013) definieren, durch die Unendlich-Norm geschieht.¹⁷ Im nachstehenden Algorithmus ist nun ersichtlich, wie diese Nachbarschaften zur Schrittlängensteuerung verwendet werden können:

Algorithmus 4.3 *Ein zulässiges „short-step“/„long-step“ Innere-Punkte-Verfahren*

Input:

- Eine Instanz des quadratischen Optimierungsproblems ($\widetilde{\text{QP}}$), welche durch $Q \in \mathbb{R}^{n \times n}$ (symmetrisch und positiv semidefinit), $c \in \mathbb{R}^n$, $B \in \mathbb{R}^{q \times n}$ und $\beta \in \mathbb{R}^q$ beschrieben wird.
- Eine Nachbarschaftsmenge $\mathcal{N} \in \{\mathcal{N}_2(\theta) \mid \theta \in (0, 1)\} \cup \{\mathcal{N}_S(\psi) \mid \psi \in (0, 1)\}$.
- Ein Punkt $(x^0, \lambda^0, \gamma^0) \in \mathcal{N}$.
- σ_{\min} und σ_{\max} mit $0 < \sigma_{\min} < \sigma_{\max} < 1$.
- ε mit $\varepsilon > 0$.

Output:

- Ein approximativer KKT-Punkt $(\hat{x}, \lambda, \gamma)$ von ($\widetilde{\text{QP}}$) (womit \hat{x} ein approximatives globales Minimum ist).
-

- 1: Setze $K := 0$.
- 2: Setze $\mu^K := \frac{1}{n}(x^K)^\top \lambda^K$.
- 3: **if** $\mu^K \leq \varepsilon$ **then**
- 4: **return** $(x^K, \lambda^K, \gamma^K)$
- 5: **end if**

¹⁷In der Literatur (siehe z.B. Cornuéjols und Tütüncü, 2007) wird anstatt der symmetrischen Nachbarschaft oftmals die einseitige Nachbarschaft $\mathcal{N}_{-\infty}(\psi)$ verwendet, deren Punkte nur $x_i\lambda_i \geq \psi\mu$ erfüllen müssen. Gondzio (2013) verweist jedoch auf theoretische Vorteile der symmetrischen Nachbarschaft.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

6: Wähle $\sigma^K \in [\sigma_{\min}, \sigma_{\max}]$ und berechne $(\Delta x^K, \Delta \lambda^K, \Delta \gamma^K)$ als Lösung von (4.18).

7: Setze

$$(x^{K+1}, \lambda^{K+1}, \gamma^{K+1}) := (x^K, \lambda^K, \gamma^K) + t^K (\Delta x^K, \Delta \lambda^K, \Delta \gamma^K)$$

mit einem **größtmöglichen** $t^K \in (0, 1]$, sodass

$$(x^{K+1}, \lambda^{K+1}, \gamma^{K+1}) \in \mathcal{N}$$

gilt.

8: Setze $K := K + 1$.

9: **goto** Zeile 2.

Wird im obigen Algorithmus die euklidische Nachbarschaft $\mathcal{N}_2(\theta)$ gewählt, so spricht man von einem „short-step“-Verfahren. Die symmetrische Nachbarschaft $\mathcal{N}_S(\psi)$ führt zu einer „long-step“-Methode. Diese Bezeichnungen lassen sich darauf zurückführen, dass $\mathcal{N}_S(\psi)$ nahezu alle Punkte aus \mathcal{F}° beinhaltet, falls ψ hinreichend klein gewählt wird. Diese Eigenschaft ist hingegen für $\mathcal{N}_2(\theta)$ selbst dann nicht gewährleistet, wenn θ nahe 1 ist (siehe Nocedal und Wright, 2006). Darüber hinaus gilt folgende Aussage:¹⁸

Lemma 4.1.6. *Für $\psi \leq 1 - \theta$ ist $\mathcal{N}_2(\theta) \subseteq \mathcal{N}_S(\psi)$.*

Beweis. Sei $(x, \lambda, \gamma) \in \mathcal{N}_2(\theta)$. Dann gilt

$$\|X\lambda e - \mu e\|_2^2 = \sum_{j=1}^n (x_j \lambda_j - \mu)^2 = (x_i \lambda_i - \mu)^2 + \sum_{\substack{j=1 \\ j \neq i}}^n (x_j \lambda_j - \mu)^2 \leq \theta^2 \mu^2. \quad (4.20)$$

Laut Annahme ist $\psi \leq 1 - \theta$. Daraus folgt wegen $(1 + \theta) \cdot (1 - \theta) = 1 - \theta^2 \leq 1$, dass

$$(1 + \theta) \leq \frac{1}{1 - \theta} \leq \frac{1}{\psi}.$$

Nach Umformulierung erhalten wir deshalb aus (4.20) die Ungleichungskette

$$\psi \mu \leq (1 - \theta) \mu \leq x_i \lambda_i \leq (1 + \theta) \mu \leq \frac{1}{\psi} \mu$$

für jedes beliebige $i \in \{1, \dots, n\}$, womit $(x, \lambda, \gamma) \in \mathcal{N}_S(\psi)$ und insgesamt $\mathcal{N}_2(\theta) \subseteq \mathcal{N}_S(\psi)$ resultiert. \square

¹⁸Vgl. Cornuéjols und Tütüncü (2007), welche eine analoge Aussage für $\mathcal{N}_2(\theta)$ und die einseitige Nachbarschaft $\mathcal{N}_{-\infty}(\psi)$ formulieren.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

4.1.3.2. Konvergenz der Innere-Punkte-Verfahren mit Nachbarschaften

Wir können nun unter Verwendung dieser Nachbarschaften Konvergenzeigenschaften angeben. Für die nachstehenden Resultate, welche von Gondzio (2013) für das Problem der quadratischen Optimierung hergeleitet werden und die polynomielle Komplexität des „short-step“ und des „long-step“ Innere-Punkte-Verfahrens bei geeigneter Parameterwahl zeigen, betrachten wir folgende Modifikation von Algorithmus 4.3:

Algorithmus 4.4 *Ein zulässiges „short-step“/„long-step“ Innere-Punkte-Verfahren*

Input:

- Eine Instanz des quadratischen Optimierungsproblems $(\widetilde{\text{QP}})$, welche durch $Q \in \mathbb{R}^{n \times n}$ (symmetrisch und positiv semidefinit), $c \in \mathbb{R}^n$, $B \in \mathbb{R}^{q \times n}$ und $\beta \in \mathbb{R}^q$ beschrieben wird.
- Eine Nachbarschaftsmenge $\mathcal{N} \in \{\mathcal{N}_2(\theta) \mid \theta \in (0, 1)\} \cup \{\mathcal{N}_S(\psi) \mid \psi \in (0, 1)\}$.
- Ein Punkt $(x^0, \lambda^0, \gamma^0) \in \mathcal{N}$.
- σ mit $0 < \sigma < 1$ und t mit $0 < t \leq 1$, sodass sämtliche Punkte $(x^{K+1}, \lambda^{K+1}, \gamma^{K+1})$ mit $K \geq 0$, die im Laufe des Algorithmus generiert werden, in \mathcal{N} liegen.
- ε mit $\varepsilon > 0$.

Output:

- Ein approximativer KKT-Punkt $(\hat{x}, \lambda, \gamma)$ von $(\widetilde{\text{QP}})$ (womit \hat{x} ein approximatives globales Minimum ist).
-

- 1: Setze $K := 0$.
- 2: Setze $\mu^K := \frac{1}{n}(x^K)^\top \lambda^K$.
- 3: **if** $\mu^K \leq \varepsilon$ **then**
- 4: **return** $(x^K, \lambda^K, \gamma^K)$
- 5: **end if**
- 6: Berechne $(\Delta x^K, \Delta \lambda^K, \Delta \gamma^K)$ als Lösung von (4.18) mit $\sigma^K = \sigma$.
- 7: Setze

$$(x^{K+1}, \lambda^{K+1}, \gamma^{K+1}) := (x^K, \lambda^K, \gamma^K) + t(\Delta x^K, \Delta \lambda^K, \Delta \gamma^K).$$
- 8: Setze $K := K + 1$.
- 9: **goto** Zeile 2.

Für diesen Algorithmus gelten nun folgende Aussagen:

Satz 4.1.7. *Es seien ein konvexes Optimierungsproblem $(\widetilde{\text{QP}})$ mit der Nachbarschaft $\mathcal{N}_2(0.1)$, $\varepsilon > 0$ sowie ein zulässiger Startpunkt $(x^0, \lambda^0, \gamma^0) \in \mathcal{N}_2(0.1)$ für Algorithmus 4.4 gegeben, sodass $\mu_0 \leq \frac{1}{\varepsilon^\kappa}$ für eine positive Konstante κ (mit $\mu_0 = \frac{1}{n}(x^0)^\top \lambda^0$) gilt. Weiters sei $\sigma = 1 - 0.1/\sqrt{n}$ und $t = 1$. Dann existiert ein $L \in \mathbb{N}$ mit $L = \mathcal{O}(\sqrt{n} \ln(1/\varepsilon))$, sodass $\mu^K \leq \varepsilon$ für alle $K \geq L$ erfüllt ist.*

Satz 4.1.8. *Es seien ein konvexes Optimierungsproblem $(\widetilde{\text{QP}})$ mit der Nachbarschaft $\mathcal{N}_S(0.5)$, $\varepsilon > 0$ sowie ein zulässiger Startpunkt $(x^0, \lambda^0, \gamma^0) \in \mathcal{N}_S(0.5)$ für Algorithmus 4.4*

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

gegeben, sodass $\mu_0 \leq \frac{1}{\varepsilon^\kappa}$ für eine positive Konstante κ (mit $\mu_0 = \frac{1}{n}(x^0)^\top \lambda^0$) gilt. Weiters sei $\sigma = 0.5$ und $t = 1/50n$. Dann existiert ein $L \in \mathbb{N}$ mit $L = \mathcal{O}(n \ln(1/\varepsilon))$, sodass $\mu^K \leq \varepsilon$ für alle $K \geq L$ erfüllt ist.

Die Herleitungen dieser beiden Aussagen von Gondzio (2013) beinhalten folgende Eigenschaften:

- Die Beweise werden auf entsprechende Überlegungen von Wright (1997) für lineare Optimierungsprobleme zurückgeführt.
- Mit der getroffenen Wahl von t und σ in Satz 4.1.7 bzw. 4.1.8 ist automatisch gewährleistet, dass $(x^{K+1}, \lambda^{K+1}, \gamma^{K+1})$ für jedes $K \geq 0$ in der entsprechenden Nachbarschaft und damit auch in der strikt zulässigen Menge liegt.
- Beide Sätze werden von Gondzio (2013) sogar unter der Annahme bewiesen, dass die dritte Gleichung in (4.18) in jeder Iteration von Algorithmus 4.4 durch

$$\Lambda^K \Delta x^K + X^K \Delta \lambda^K = -X^K \Lambda^K e + \sigma^K \mu^K e + r$$

ersetzt wird. Dabei ist r ein Fehlerterm mit $\|r\|_2 \leq 0.3 \cdot \|-X^K \Lambda^K e + \sigma^K \mu^K e\|_2$ in Satz 4.1.7 bzw. mit $\|r\|_\infty \leq 0.05 \cdot \|-X^K \Lambda^K e + \sigma^K \mu^K e\|_\infty$ in Satz 4.1.8. Dies kann in der Praxis einen großen Vorteil für die Laufzeit der Algorithmen bedeuten: Nocedal und Wright (2006) verweisen nämlich darauf, dass Innere-Punkte-Methoden im Allgemeinen weniger Iterationsschritte als die Methode der aktiven Menge benötigen, ein einzelner Iterationsschritt hingegen aufwendiger ist.¹⁹ Daher kann es sinnvoll, das Gleichungssystem (4.18) numerisch mit iterativen Methoden zu lösen (vgl. D’Apuzzo, Simone und Serafino, 2010), was besonders effizient realisierbar ist, wenn die Gleichungen nur approximativ (also bezüglich eines hinreichend kleinen Fehlers) erfüllt sein müssen. Gondzio (2013) spricht in diesem Fall von *inexakten zulässigen Innere-Punkte-Methoden* („inexact feasible interior point methods“).

Damit wissen wir nun, dass die Verwendung der restriktiveren euklidischen Nachbarschaft bessere „Worst-Case“-Laufzeiteigenschaften liefert als die „long-step“-Variante mit der symmetrischen Nachbarschaft. Gondzio (2013) verweist jedoch unter anderem darauf, dass in der Praxis für $N_2(\theta)$ auch tatsächlich ein Laufzeitverhalten im „Worst-Case“-Bereich zu beobachten ist, während $N_S(\psi)$ oftmals bessere Ergebnisse liefert und daher in praktischen Anwendungen von großer Bedeutung ist.

4.1.3.3. Ein unzulässiges Innere-Punkte-Verfahren

Wie wir im Zusammenhang mit der Methode der aktiven Menge gesehen haben, ist es einfach, für ein quadratisches Programm einen primal zulässigen Punkt zu finden oder die Unzulässigkeit festzustellen. Allerdings kann die Voraussetzung der Algorithmen 4.3 und 4.4, mit einem *strikt* zulässigen Startpunkt zu beginnen, problematisch sein.

¹⁹Diese Aussage wird auch durch unsere empirischen Untersuchungen in Kapitel 5 untermauert.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

Um dies zu umgehen, schlägt Wright (1997) für lineare Optimierungsprobleme einen „long-step“-Algorithmus vor, in dem die einseitige Nachbarschaft $\mathcal{N}_{-\infty}(\psi)$ so angepasst wird, dass mit einem *beliebigen* (also auch unzulässigen) Punkt gestartet werden kann. Wir möchten hingegen einen kurzen Überblick über die sogenannte „Predictor-Corrector“-Methode nach Mehrotra (1992) geben. Auch dieses Verfahren ermöglicht einen unzulässigen Startvektor, vermeidet jedoch die explizite Verwendung von Nachbarschaften. Die folgenden Ausführungen entsprechen der Vorgangsweise von Nocedal und Wright (2006).

Das „Predictor-Corrector“-Verfahren gliedert sich (wie der Name bereits andeutet) in zwei Schritte pro Iteration. Wir untersuchen die $(K + 1)$ -te Iteration mit gegebenem (gegebenenfalls unzulässigem) Punkt $(x^K, \lambda^K, \gamma^K)$ mit $x^K > 0$ und $\lambda^K > 0$:

Im „*Predictor*“-Schritt wird eine reine Newton-Richtung als Lösung von

$$\begin{pmatrix} Q & -I_n & -B^\top \\ B & 0 & 0 \\ \Lambda^K & X^K & 0 \end{pmatrix} \cdot \begin{pmatrix} \Delta_{pre} x^K \\ \Delta_{pre} \lambda^K \\ \Delta_{pre} \gamma^K \end{pmatrix} = \begin{pmatrix} -r_d^K \\ -r_p^K \\ -X^K \Lambda^K e \end{pmatrix} \quad (4.21)$$

berechnet. Wie wir bereits gesehen haben, zielt dieses Prozedere darauf ab, einen neuen Iterationspunkt zu erhalten, welcher (unter der theoretischen Annahme, dass die volle Schrittlänge verwendet wird) die komplementäre Schlupfbedingung c) erfüllen sollte. Da jedoch eine Taylorapproximation 1. Ordnung und damit eine Linearisierung durchgeführt wurde, gilt wegen

$$\Lambda^K \Delta_{pre} x^K + X^K \Delta_{pre} \lambda^K = -X^K \Lambda^K e,$$

dass

$$(x_i + \Delta_{pre} x_i^K) (\lambda_i + \Delta_{pre} \lambda_i^K) = \Delta_{pre} x_i^K \Delta_{pre} \lambda_i^K.$$

Um diesen Term höherer Ordnung nicht zu ignorieren, wird im „*Corrector*“-Schritt (nun wieder unter Verwendung eines zentrierenden Parameters σ^K) das Gleichungssystem (4.18) ausgehend vom zuvor berechneten Vektor $(\Delta_{pre} x^K, \Delta_{pre} \lambda^K, \Delta_{pre} \gamma^K)$ betrachtet. Es ist also

$$\begin{pmatrix} Q & -I_n & -B^\top \\ B & 0 & 0 \\ \Lambda^K & X^K & 0 \end{pmatrix} \cdot \begin{pmatrix} \Delta_{cor} x^K \\ \Delta_{cor} \lambda^K \\ \Delta_{cor} \gamma^K \end{pmatrix} =: \begin{pmatrix} 0 \\ 0 \\ -\Delta_{pre} X^K \Delta_{pre} \Lambda^K e + \sigma^K \mu^K e \end{pmatrix} \quad (4.22)$$

mit $\Delta_{pre} X^K := \text{diag}(\Delta_{pre} x_1^K, \dots, \Delta_{pre} x_n^K)$ und $\Delta_{pre} \Lambda^K := \text{diag}(\Delta_{pre} \lambda_1^K, \dots, \Delta_{pre} \lambda_n^K)$ zu lösen. Aus der Kombination der beiden berechneten Richtungen ergibt sich insgesamt der Vektor

$$\Delta x^K := \Delta_{pre} x^K + \Delta_{cor} x^K,$$

womit offensichtlich im zweiten Schritt das System (4.22) auch direkt durch

$$\begin{pmatrix} Q & -I_n & -B^\top \\ B & 0 & 0 \\ \Lambda^K & X^K & 0 \end{pmatrix} \cdot \begin{pmatrix} \Delta x^K \\ \Delta \lambda^K \\ \Delta \gamma^K \end{pmatrix} =: \begin{pmatrix} -r_d^K \\ -r_p^K \\ -X^K \Lambda^K e - \Delta_{pre} X^K \Delta_{pre} \Lambda^K e + \sigma^K \mu^K e \end{pmatrix} \quad (4.23)$$

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

ersetzt werden kann. Da sich die beiden relevanten Gleichungssysteme (4.21) und (4.23) nur in der rechten Seite unterscheiden, ist der zusätzliche Aufwand für die Lösung des zweiten Systems vernachlässigbar.

Die Aufteilung in einen „Predictor“- und einen „Corrector“-Teil hat darüber hinaus den Vorteil, den zentrierenden Parameter σ^K nach dem „Predictor“-Schritt plausibel wählen zu können. Dafür sei

$$t_{pre}^K := \max \{t \in (0, 1] \mid x^K + t\Delta_{pre}x^K \geq 0 \text{ und } \lambda^K + t\Delta_{pre}\lambda^K \geq 0\}$$

und

$$\mu_{pre}^K := \frac{1}{n} (x^K + t_{pre}^K \Delta_{pre} x^K)^\top (\lambda^K + t_{pre}^K \Delta_{pre} \lambda^K)$$

das Dualitätsmaß des *theoretisch* resultierenden Punktes mit Schrittlänge t_{pre}^K . Grundsätzlich sind kleine Werte für σ^K zu bevorzugen, da das Ziel eine Reduktion des Dualitätsmaßes $\mu^K = \frac{1}{n} (x^K)^\top \lambda^K$ ist. Nur wenn die reine Newton-Richtung zu einer sehr kleinen Schrittweite bzw. zu einer sehr geringen Reduktion des Dualitätsmaßes führen würde, ist es sinnvoll, einen stark zentrierenden Schritt (mit $\sigma^K \approx 1$) durchzuführen, um in der nächsten Iteration (hoffentlich) das Dualitätsmaß stärker verringern zu können. Basierend auf diesen Überlegungen schlagen Nocedal und Wright (2006)

$$\sigma^K := (\mu_{pre}^K / \mu^K)^3$$

vor.²⁰

Als Schrittweite für unseren neuen Iterationspunkt wählen wir schlussendlich

$$t^K := \max \{t \in (0, 1] \mid x^K + t\Delta x^K \geq (1 - l^K)x^K \text{ und } \lambda^K + t\Delta\lambda^K \geq (1 - l^K)\lambda^K\}$$

für ein passendes $l^K \in (0, 1)$, um zu gewährleisten, dass die Nichtnegativitätsbedingungen strikt erfüllt bleiben. Damit gilt wegen (4.23)

$$\begin{pmatrix} r_d^{K+1} \\ r_p^{K+1} \end{pmatrix} = (1 - t^K) \begin{pmatrix} r_d^K \\ r_p^K \end{pmatrix},$$

was zu einer entsprechenden Reduktion der Residuen im Laufe des Algorithmus führt. Dabei sollte l^K gegen 1 gehen, wenn sich der Algorithmus der gesuchten Lösung nähert, damit die komplementäre Schlupfbedingung approximativ erfüllbar ist.

Wir können nun die obigen Ideen in nachstehendem Pseudoalgorithmus zusammenfassen, welcher jeden beliebigen Startpunkt mit $x^0 > 0$ und $\lambda^0 > 0$ akzeptiert:

²⁰Nocedal und Wright (2006) weisen darauf hin, dass diese Wahl in der Praxis gute Ergebnisse liefert, jedoch auf keiner analytischen Herleitung beruht.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

Algorithmus 4.5 *Ein unzulässiges „Predictor-Corrector“-Verfahren nach Mehrotra (1992)*

Input:

- Eine Instanz des quadratischen Optimierungsproblems ($\widetilde{\text{QP}}$), welche durch $Q \in \mathbb{R}^{n \times n}$ (symmetrisch und positiv semidefinit), $c \in \mathbb{R}^n$, $B \in \mathbb{R}^{q \times n}$ und $\beta \in \mathbb{R}^q$ beschrieben wird.
- Ein Punkt $(x^0, \lambda^0, \gamma^0) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^q$ mit $x^0 > 0$ und $\lambda^0 > 0$.
- Eine Folge $\{l^K\}_{K=0,1,\dots}$ mit $l^K \in (0, 1)$ und $l^K \xrightarrow{K \rightarrow \infty} 1$.
- $\varepsilon_p > 0$, $\varepsilon_d > 0$ und $\varepsilon_s > 0$.

Output:

- Ein approximativer KKT-Punkt $(\hat{x}, \lambda, \gamma)$ von ($\widetilde{\text{QP}}$) (womit \hat{x} ein approximatives globales Minimum ist).
-

- 1: Setze $K := 0$.
- 2: Setze $r_p^K := Bx^K - \beta$, $r_d^K := Qx^K + c - \lambda^K - B^\top \gamma^K$ und $\mu^K := \frac{1}{n}(x^K)^\top \lambda^K$.
- 3: **if** $\mu^K \leq \varepsilon_s$ **and** $\|r_p^K\|_2 \leq \varepsilon_p$ **and** $\|r_d^K\|_2 \leq \varepsilon_d$ **then**
- 4: **return** $(x^K, \lambda^K, \gamma^K)$
- 5: **end if**
- 6: Berechne $(\Delta_{pre}x^K, \Delta_{pre}\lambda^K, \Delta_{pre}\gamma^K)$ als Lösung von (4.21).
- 7: Berechne $t_{pre}^K := \max \{t \in (0, 1] \mid x^K + t\Delta_{pre}x^K \geq 0 \text{ und } \lambda^K + t\Delta_{pre}\lambda^K \geq 0\}$.
- 8: Setze $\sigma^K := (\mu_{pre}^K/\mu^K)^3$ mit $\mu_{pre}^K = \frac{1}{n}(x^K + t_{pre}^K\Delta_{pre}x^K)^\top (\lambda^K + t_{pre}^K\Delta_{pre}\lambda^K)$.
- 9: Berechne $(\Delta x^K, \Delta \lambda^K, \Delta \gamma^K)$ als Lösung von (4.23).
- 10: Berechne

$$t^K := \max \{t \in (0, 1] \mid x^K + t\Delta x^K \geq (1 - l^K)x^K \text{ und } \lambda^K + t\Delta \lambda^K \geq (1 - l^K)\lambda^K\}.$$

- 11: Setze

$$(x^{K+1}, \lambda^{K+1}, \gamma^{K+1}) := (x^K, \lambda^K, \gamma^K) + t^K(\Delta x^K, \Delta \lambda^K, \Delta \gamma^K).$$

- 12: Setze $K := K + 1$.

- 13: **goto** Zeile 2.

Es sei noch erwähnt, dass es laut Nocedal und Wright (2006) für lineare Optimierungsprobleme möglich ist, verschieden große Schrittweiten für die primalen sowie die dualen Variablen zu wählen, was in der Praxis oftmals zu einer Verringerung der benötigten Anzahl von Iterationen führt. Würde man jedoch für ein quadratisches Programm eine primale Schrittweite t_p durch $x^K + t_p\Delta x^K > 0$ und eine davon unabhängige duale Schrittweite t_d mithilfe der Bedingung $\lambda^K + t_d\Delta \lambda^K > 0$ ohne weitere Einschränkungen bestimmen, so könnte dies aufgrund des quadratischen Terms dazu führen, dass $r_d^{K+1} > r_d^K$ ist. Aus diesem Grund haben wir uns auf gleich große Schrittweiten für alle Variablen beschränkt. Curtis und Nocedal (2007) behandeln allerdings eine Möglichkeit, ungleiche Schrittweiten zu realisieren, und haben damit auch für quadratische Optimierungsprobleme eine Reduktion der Iterationszahl beobachtet.

4.2. Sequentielle Quadratische Optimierung

Nachdem wir nun Möglichkeiten zur Lösung konvexer quadratischer Optimierungsprobleme kennengelernt haben, möchten wir als Nächstes allgemeinere nichtlineare Programme untersuchen. Wir werden zu Beginn einige Aussagen für Probleme der Form (NLP) angeben, uns in weiterer Folge jedoch auf die kleinere Problemklasse

$$\begin{aligned} (\widetilde{\text{NLP}}) \quad & \min_{x \in \mathbb{R}^n} f(x) \\ & s.t. \quad Ax \geq \alpha \\ & \quad \quad Bx = \beta \end{aligned}$$

mit linearen Restriktionen beschränken, da die für uns relevanten Probleme (KQ1) und (KQ2) (sowie auch (LS $_{\theta}^c$) aus Abschnitt 3.3.1) als $(\widetilde{\text{NLP}})$ darstellbar sind.

Alle Funktionen, die im Folgenden auftreten, seien sooft stetig differenzierbar wie nötig.

4.2.1. Behandlung von Problemen mit Gleichheitsrestriktionen

Die Theorie dieses Abschnittes stammt aus Geiger und Kanzow (2002) sowie Nocedal und Wright (2006).

Wir betrachten einführend folgende nichtlineare Optimierungsprobleme:

$$\begin{aligned} (\widehat{\text{NLP}}) \quad & \min_{x \in \mathbb{R}^n} f(x) \\ & s.t. \quad h_k(x) = 0, \quad k = 1, \dots, q \end{aligned}$$

Da $(\widehat{\text{NLP}})$ keine Ungleichheitsrestriktionen beinhaltet, sind die entsprechenden KKT-Bedingungen besonders einfach anzugeben. Ein KKT-Punkt $(x^*, \gamma) \in \mathbb{R}^n \times \mathbb{R}^q$ erfüllt folgende Voraussetzungen:

$$\begin{aligned} [\widehat{\text{NLP}}\text{-i}] \quad & \nabla_x \widehat{\mathcal{L}}(x, \gamma)|_{x=x^*} = 0, \\ [\widehat{\text{NLP}}\text{-ii}] \quad & h(x^*) = 0. \end{aligned}$$

Dabei sei $h(x) := (h_1(x), \dots, h_q(x))^{\top}$ und $\widehat{\mathcal{L}}(x, \gamma) := f(x) - \sum_{k=1}^q \gamma_k h_k(x) = f(x) - h(x)^{\top} \gamma$ die Lagrangefunktion von $(\widehat{\text{NLP}})$. Offensichtlich lässt sich nun ein KKT-Punkt ermitteln, indem man auf die Funktion

$$\begin{aligned} F : \mathbb{R}^{n+q} &\rightarrow \mathbb{R}^{n+q} \\ (x, \gamma) &\mapsto \begin{pmatrix} \nabla_x \widehat{\mathcal{L}}(x, \gamma) \\ h(x) \end{pmatrix} \end{aligned}$$

das in Appendix A.6 beschriebene Newton-Verfahren (Algorithmus A.2) anwendet. Die resultierende Methode zur Lösung gleichheitsrestringierter Optimierungsprobleme wird von

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

Geiger und Kanzow (2002) auch *Lagrange-Newton-Verfahren* genannt. Für die Jacobi-Matrix der Funktion F gilt nun

$$\mathcal{J}_F(x, \gamma) = \begin{pmatrix} \nabla_{xx}^2 \widehat{\mathcal{L}}(x, \gamma) & -\mathcal{J}_h(x)^\top \\ \mathcal{J}_h(x) & 0 \end{pmatrix},$$

wobei $\mathcal{J}_h(x)$ die Jacobi-Matrix von h in x darstellt. Somit ist in der $(K + 1)$ -ten Iteration (ausgehend vom Punkt (x^K, γ^K)) das Gleichungssystem

$$\begin{pmatrix} \nabla_{xx}^2 \widehat{\mathcal{L}}(x, \gamma^K)|_{x=x^K} & -\mathcal{J}_h(x^K)^\top \\ \mathcal{J}_h(x^K) & 0 \end{pmatrix} \begin{pmatrix} \Delta x^K \\ \Delta \gamma^K \end{pmatrix} = \begin{pmatrix} -\nabla_x \widehat{\mathcal{L}}(x, \gamma^K)|_{x=x^K} \\ -h(x^K) \end{pmatrix} \quad (4.24)$$

zu lösen ist. Dafür ist der nachstehende Satz relevant:

Satz 4.2.1. (x^*, γ) sei ein LICQ-regulärer KKT-Punkt für $(\widehat{\text{NLP}})$, in dem die hinreichenden KKT-Bedingungen 2. Ordnung aus Satz 1.4.4 erfüllt sind. Dann ist die Jacobi-Matrix $\mathcal{J}_F(x^*, \gamma)$ regulär.

Beweis. (Vgl. Geiger und Kanzow, 2002.)

Die Voraussetzungen des Satzes können wegen des Fehlens von Ungleichheitsrestriktionen auch folgendermaßen formuliert werden:

- i) Die Gradientenvektoren $\nabla h_k(x^*)$ mit $k \in \{1, \dots, q\}$ sind linear unabhängig.
- ii) Für $y \in \mathbb{R}^n$ mit $y \neq 0$ und $\nabla h_k(x^*)^\top y = 0, \forall k = 1, \dots, q$ ist $y^\top \nabla_{xx}^2 \widehat{\mathcal{L}}(x, \gamma)|_{x=x^*} y > 0$.

Bedingung ii) ist dabei äquivalent zu der Annahme aus Satz 1.4.4, dass die Matrix $B_h(x^*)^\top \nabla_{xx}^2 \widehat{\mathcal{L}}(x, \gamma)|_{x=x^*} B_h(x^*)$ positiv definit ist für eine Basis $B_h(x^*)$ des Kerns der Jacobi-Matrix $\mathcal{J}_h(x^*)$.

Wir betrachten nun das homogene Gleichungssystem

$$\mathcal{J}_F(x^*, \gamma) \begin{pmatrix} \hat{x} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} \nabla_{xx}^2 \widehat{\mathcal{L}}(x, \gamma)|_{x=x^*} & -\mathcal{J}_h(x^*)^\top \\ \mathcal{J}_h(x^*) & 0 \end{pmatrix} \begin{pmatrix} \hat{x} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

welches wir auch wie folgt schreiben können:

$$\nabla_{xx}^2 \widehat{\mathcal{L}}(x, \gamma)|_{x=x^*} \hat{x} - \sum_{k=1}^q \hat{\gamma}_k \nabla h_k(x^*) = 0 \quad (4.25)$$

$$\nabla h_k(x^*)^\top \hat{x} = 0, \forall k = 1, \dots, q \quad (4.26)$$

Multiplizieren wir nun Gleichung (4.25) von links mit \hat{x}^\top und setzen (4.26) ein, so folgt $\hat{x}^\top \nabla_{xx}^2 \widehat{\mathcal{L}}(x, \gamma)|_{x=x^*} \hat{x} = 0$ und damit wegen (4.26) sowie Bedingung ii) sofort $\hat{x} = 0$. Schlussendlich liefert (4.25) wegen $\hat{x} = 0$ und Bedingung i) auch $\hat{\gamma} = 0$, womit die Matrix $\mathcal{J}_F(x^*, \gamma)$ regulär ist. \square

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

Unter den in obigem Satz getroffenen Annahmen ist also Satz A.6.1 anwendbar, welcher besagt, dass (4.24) in einer *Umgebung* von x^* eindeutig lösbar ist. Darüber hinaus ist das Lagrange-Newton-Verfahren *lokal* konvergent mit superlinearer bzw. (unter der zusätzlichen Voraussetzung, dass \mathcal{J}_F lokal Lipschitz-stetig ist) quadratischer Konvergenzrate.

Wir möchten nun noch eine andere Möglichkeit zur Herleitung des Gleichungssystems (4.24) kennenlernen, um davon ausgehend eine Erweiterung auf Probleme mit Ungleichheitsrestriktionen durchzuführen. Dazu gehen wir davon aus, dass wir in der K -ten Iteration des Lagrange-Newton-Verfahrens den neuen Iterationspunkt (x^K, γ^K) berechnet haben, und betrachten das folgende Optimierungsproblem, welches äquivalent zu $(\widehat{\text{NLP}})$ ist:

$$\begin{aligned} \min_{\Delta x^K \in \mathbb{R}^n} \quad & \widehat{\mathcal{L}}(x^K + \Delta x^K, \gamma^K) \\ \text{s.t.} \quad & h(x^K + \Delta x^K) = 0 \end{aligned}$$

Die Äquivalenz folgt aus der Reparametrisierung $x = x^K + \Delta x^K$ und der Tatsache, dass $\widehat{\mathcal{L}}(x, \gamma^K) = f(x) - h(x)^\top \gamma^K = f(x)$ für alle zulässigen Punkte x gilt.

Als Nächstes führen wir eine lineare Taylorapproximation der Gleichheitsrestriktionen durch und verwenden als Zulässigkeitsmenge all jene Punkte, in denen ebendiese Approximation gleich 0 ist:

$$h(x^K + \Delta x^K) \approx h(x^K) + \mathcal{J}_h(x^K) \Delta x^K = 0 \quad (4.27)$$

Für die Zielfunktion führen wir dagegen eine Taylorapproximation 2. Ordnung bezüglich des ersten Arguments durch. Diese lässt sich für all jene Punkte, die (4.27) erfüllen, folgendermaßen schreiben:

$$\begin{aligned} \widehat{\mathcal{L}}(x^K + \Delta x^K, \gamma^K) &\approx \widehat{\mathcal{L}}(x^K, \gamma^K) + \nabla_x \widehat{\mathcal{L}}(x, \gamma^K) \Big|_{x=x^K}^\top \Delta x^K \\ &\quad + \frac{1}{2} (\Delta x^K)^\top \nabla_{xx}^2 \widehat{\mathcal{L}}(x, \gamma^K) \Big|_{x=x^K} \Delta x^K \\ &= f(x^K) - h(x^K)^\top \gamma^K + \nabla f(x^K)^\top \Delta x^K - (\Delta x^K)^\top \mathcal{J}_h(x^K)^\top \gamma^K \\ &\quad + \frac{1}{2} (\Delta x^K)^\top \nabla_{xx}^2 \widehat{\mathcal{L}}(x, \gamma^K) \Big|_{x=x^K} \Delta x^K \\ &= f(x^K) + \nabla f(x^K)^\top \Delta x^K + \frac{1}{2} (\Delta x^K)^\top \nabla_{xx}^2 \widehat{\mathcal{L}}(x, \gamma^K) \Big|_{x=x^K} \Delta x^K \end{aligned}$$

Insgesamt erhalten wir (unter Berücksichtigung, dass $f(x^K)$ konstant und somit für die Optimierung irrelevant ist) ein quadratisches Optimierungsproblem als Approximation von $(\widehat{\text{NLP}})$:

$$\begin{aligned} (\widehat{\text{NLP}}^K) \quad & \min_{\Delta x^K \in \mathbb{R}^n} \quad \nabla f(x^K)^\top \Delta x^K + \frac{1}{2} (\Delta x^K)^\top \nabla_{xx}^2 \widehat{\mathcal{L}}(x, \gamma^K) \Big|_{x=x^K} \Delta x^K \\ & \text{s.t.} \quad h(x^K) + \mathcal{J}_h(x^K) \Delta x^K = 0 \end{aligned}$$

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

Nun untersuchen wir einen KKT-Punkt $(\Delta x^K, \tilde{\gamma})$ dieses Problems. Verwenden wir für die Lagrange-Multiplikatoren die Reparametrisierung $\tilde{\gamma} = \gamma^K + \Delta\gamma^K$, so erhalten wir für die KKT-Bedingung an den Gradienten der Lagrangefunktion die nachstehende Gleichung:

$$\begin{aligned} \nabla f(x^K) + \nabla_{xx}^2 \widehat{\mathcal{L}}(x, \gamma^K)|_{x=x^K} \Delta x^K - \mathcal{J}_h(x^K)^\top \tilde{\gamma} &= 0 \\ \Leftrightarrow \\ \nabla_{xx}^2 \widehat{\mathcal{L}}(x, \gamma^K)|_{x=x^K} \Delta x^K - \mathcal{J}_h(x^K)^\top \Delta\gamma^K &= -\nabla_x \mathcal{L}(x, \gamma^K)|_{x=x^K} \end{aligned}$$

Dies entspricht zusammen mit der Zulässigkeitsbedingung

$$\mathcal{J}_h(x^K) \Delta x^K = -h(x^K)$$

jedoch genau wieder dem linearen Gleichungssystem (4.24). Wir haben damit gezeigt, dass in jeder Iteration des Lagrange-Newton-Verfahrens anstatt einer Lösung von (4.24) auch ein KKT-Punkt des quadratischen Problems $(\widehat{\text{NLP}}^K)$ berechnet werden kann. Als neuer Iterationspunkt ergibt sich $x^{K+1} := x^K + \Delta x^K$ und $\gamma^{K+1} := \tilde{\gamma}$.

4.2.2. Eine lokale SQP-Methode für Probleme mit Ungleichheitsrestriktionen

Um nun nichtlineare Programme (NLP) behandeln zu können, die auch Ungleichheitsrestriktionen beinhalten, ist es naheliegend, das im letzten Unterabschnitt eingeführte Hilfsproblem $(\widehat{\text{NLP}}^K)$ entsprechend zu erweitern. Dies geschieht durch Hinzunahme der Ungleichheitsrestriktionen, welche so wie die Gleichheitsrestriktionen linearisiert werden. Die Hesse-Matrix $\nabla_{xx}^2 \mathcal{L}(x, \lambda^K, \gamma^K)|_{x=x^K}$ der Lagrangefunktion

$$\mathcal{L}(x, \lambda, \gamma) := f(x) - g(x)^\top \lambda - h(x)^\top \gamma$$

mit $g(x) := (g_1(x), \dots, g_p(x))^\top$, welche in der Zielfunktion des erweiterten, quadratischen Hilfsproblems der $(K+1)$ -ten Iteration auftritt, wird dabei jedoch häufig durch eine adäquate approximierende symmetrische Matrix L_K ersetzt. Das Resultat dieser Überlegungen ist ein lokales SQP-Verfahren:

Algorithmus 4.6 Ein lokales SQP-Verfahren

Input:

- Eine Instanz des nichtlinearen Optimierungsproblems (NLP), welche durch $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$, $g(x) : \mathbb{R}^n \rightarrow \mathbb{R}^p$ und $h(x) : \mathbb{R}^n \rightarrow \mathbb{R}^q$ beschrieben wird.
- Ein Punkt $x^0 \in \mathbb{R}^n$.
- Eine Berechnungsvorschrift zur Ermittlung symmetrischer Matrizen $L_K \in \mathbb{R}^{n \times n}$, $K \in \{0, 1, \dots\}$.

Output:

- Ein KKT-Punkt (x^*, λ, γ) von (NLP).
-

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

- 1: Setze $K := 0$.
- 2: Ermittle $L_K \in \mathbb{R}^{n \times n}$.
- 3: Berechne einen KKT-Punkt $(\Delta x^K, \lambda^{K+1}, \gamma^{K+1})$ des folgenden quadratischen Programms:

$$\begin{aligned}
 (\text{NLP}^K) \quad & \min_{\Delta x^K \in \mathbb{R}^n} \quad \nabla f(x^K)^\top \Delta x^K + \frac{1}{2} (\Delta x^K)^\top L_K \Delta x^K \\
 & \text{s.t.} \quad g(x^K) + \mathcal{J}_g(x^K) \Delta x^K \geq 0 \\
 & \quad \quad h(x^K) + \mathcal{J}_h(x^K) \Delta x^K = 0
 \end{aligned}$$

- 4: Setze $x^{K+1} := x^K + \Delta x^K$.
- 5: **if** $(x^{K+1}, \lambda^{K+1}, \gamma^{K+1})$ ist ein KKT-Punkt von (NLP). **then**
- 6: **return** $(x^{K+1}, \lambda^{K+1}, \gamma^{K+1})$
- 7: **end if**
- 8: Setze $K := K + 1$.
- 9: **goto** Zeile 2.

Es ist zu beachten, dass diese Methode nicht unbedingt Startwerte λ^0 und γ^0 benötigt. In den einzelnen Iterationen erfolgt nämlich kein Update der Lagrange-Multiplikatoren durch $(\lambda^{K+1}, \gamma^{K+1}) = (\lambda^K, \gamma^K) + (\Delta \lambda^K, \Delta \gamma^K)$ wie beispielsweise in den Innere-Punkte-Verfahren für (QP), sondern eine gänzlich neue Berechnung. Falls jedoch L_K als die Hesse-Matrix der Lagrangefunktion genommen wird, sind im Allgemeinen Startpunkte λ^0 und γ^0 zur Ermittlung von L_0 notwendig.

Im obigen Algorithmus sollte eigentlich nicht nur ein KKT-Punkt, sondern ein Optimum von (NLP^K) berechnet werden. Ist (NLP) ein konvexes Optimierungsproblem der Form (KP), so ist dessen Lagrangefunktion als Summe konvexer Funktionen ebenfalls konvex und die Hesse-Matrix positiv semidefinit. Mit der Wahl $L_K = \nabla_{xx}^2 \mathcal{L}(x, \lambda^K, \gamma^K)|_{x=x^K}$ erhalten wir damit auch konvexe quadratische Probleme (NLP^K) , womit die Ermittlung eines KKT-Punktes in Zeile 3 des Algorithmus äquivalent zur Bestimmung eines Optimums ist. Für allgemeine nichtlineare Probleme ist diese Matrix hingegen nicht zwangsweise auf dem gesamten \mathbb{R}^n positiv semidefinit. Wir werden im Anschluss andere Möglichkeiten für die Wahl von L_K betrachten, sodass ebendiese Matrizen (mit einer positiv definiten Startmatrix L_0) stets positiv definit bleiben.

Wir möchten nun ein (zu Satz A.6.1 analoges) theoretisches Konvergenzresultat aus Geiger und Kanzow (2002) für Algorithmus 4.6 (mit geringfügigen Modifikationen) anführen, welches auch die Bezeichnung als „lokales SQP-Verfahren“ erklärt:

Satz 4.2.2. *(Konvergenz des lokalen SQP-Verfahrens)*

Es sei ein nichtlineares Optimierungsproblem (NLP) gegeben, auf das Algorithmus 4.6 mit $L_K = \nabla_{xx}^2 \mathcal{L}(x, \lambda^K, \gamma^K)|_{x=x^K}$ angewendet wird. Gibt es in der $(K + 1)$ -ten Iteration für (NLP^K) mehrere KKT-Punkte, so wähle man denjenigen Punkt $(\Delta x^K, \lambda^{K+1}, \gamma^{K+1})$, welcher $\|(x^K + \Delta x^K, \lambda^{K+1}, \gamma^{K+1}) - (x^K, \lambda^K, \gamma^K)\|_2$ minimiert.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

Weiters sei (x^*, λ, γ) ein LICQ-regulärer KKT-Punkt für (NLP), in dem die hinreichenden KKT-Bedingungen 2. Ordnung aus Satz 1.4.4 erfüllt sind. Dann existiert ein Wert $\varepsilon > 0$, sodass für jeden Startpunkt $(x^0, \lambda^0, \gamma^0) \in B_{(x^*, \lambda, \gamma)}(\varepsilon)$ folgende Aussagen gelten:

- i) Die von Algorithmus 4.6 erzeugte Folge $\{(x^K, \lambda^K, \gamma^K)\}_{K=0,1,\dots}$ ist wohldefiniert und konvergiert gegen (x^*, λ, γ) mit superlinearer Konvergenzrate.
- ii) Falls $\nabla^2 f(x)$, $\nabla^2 g_i(x)$, $\forall i = 1, \dots, p$ und $\nabla^2 h_k(x)$, $\forall k = 1, \dots, q$ lokal Lipschitz-stetig sind, so ist die Konvergenzrate von Algorithmus 4.6 quadratisch.

Beweis. Siehe Geiger und Kanzow (2002). (Vgl. Satz A.6.1 für die Definition einer superlinearen bzw. quadratischen Konvergenzrate.) \square

Im Zusammenhang mit diesen wünschenswerten Eigenschaften ist jedoch zu beachten, dass (wie Geiger und Kanzow (2002) anmerken) die Norm-Einschränkung an die KKT-Punkte von (NLP^K) in der Praxis schwer umsetzbar ist. Des Weiteren sollte ein praktisches SQP-Verfahren nicht nur lokal konvergieren, sondern auch mit beliebigen Punkten x^0 starten können. Wir betrachten im Folgenden eine solche Globalisierung von Algorithmus 4.6.

4.2.3. Eine globale SQP-Methode für Probleme mit Ungleichheitsrestriktionen

Bei jeder Globalisierung des lokalen SQP-Algorithmus ist von großer Bedeutung, dass für ein beliebiges nichtlineares Programm mit nichtleerer Zulässigkeitsmenge nicht zwangsläufig garantiert ist, dass auch die im lokalen SQP-Algorithmus auftretenden quadratischen Probleme (NLP^K) zulässige Punkte besitzen. Ein Beispiel für diesen Fall ist in Geiger und Kanzow (2002) angegeben. Grund dafür ist die durchgeführte Linearisierung der Restriktionen. Da in unseren Anwendungen ohnehin nur Optimierungsprogramme mit linearen Restriktionen eine Rolle spielen, für die das soeben beschriebene Problem nicht auftreten kann, möchten wir uns ab hier auf $(\widetilde{\text{NLP}})$ beschränken.

Die folgende Möglichkeit zur Globalisierung basiert auf den Ausführungen in Geiger und Kanzow (2002). Wir übertragen jedoch die dort beschriebene Vorgangsweise, welche für allgemeine nichtlineare Probleme anwendbar ist, auf unsere Problemklasse mit linearen Restriktionen. Wie wir sehen werden, stellt dies (nicht nur im Hinblick auf die Vermeidung unzulässiger quadratischer Teilprobleme) eine signifikante Vereinfachung dar.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

Betrachten wir $(\widetilde{\text{NLP}})$, so ergibt sich in Zeile 3 von Algorithmus 4.6 das folgende quadratische Problem:

$$\begin{aligned} (\widetilde{\text{NLP}}^K) \quad & \min_{\Delta x^K \in \mathbb{R}^n} \quad \nabla f(x^K)^\top \Delta x^K + \frac{1}{2} (\Delta x^K)^\top L_K \Delta x^K \\ & \text{s.t.} \quad (Ax^K - \alpha) + A\Delta x^K \geq 0 \\ & \quad \quad (Bx^K - \beta) + B\Delta x^K = 0 \end{aligned}$$

Ein KKT-Punkt $(\Delta x^K, \lambda^{K+1}, \gamma^{K+1})$ von $(\widetilde{\text{NLP}}^K)$ erfüllt folgende Bedingungen, welche wir bereits in Abschnitt 4.1.2 für quadratische Programme der Form (QP) als [QP-i], [QP-ii] und [QP-iii] formuliert haben:

$$\begin{aligned} [\widetilde{\text{NLP}}^K\text{-i}] \quad & (Ax^K - \alpha) + A\Delta x^K \geq 0, \quad (Bx^K - \beta) + B\Delta x^K = 0, \\ [\widetilde{\text{NLP}}^K\text{-ii}] \quad & L_K \Delta x^K + \nabla f(x^K) - A^\top \lambda^{K+1} - B^\top \gamma^{K+1} = 0, \\ [\widetilde{\text{NLP}}^K\text{-iii}] \quad & \lambda^{K+1} \geq 0, \quad (\lambda^{K+1})^\top \cdot (Ax^K - \alpha + A\Delta x^K) = 0. \end{aligned}$$

Tritt nun der Fall ein, dass $\Delta x^K = 0$ das Problem $(\widetilde{\text{NLP}}^K)$ löst, so ist unmittelbar ersichtlich, dass $(x^{K+1}, \lambda^{K+1}, \gamma^{K+1}) := (x^K, \lambda^{K+1}, \gamma^{K+1})$ auch ein KKT-Punkt des ursprünglichen Problems $(\widetilde{\text{NLP}})$ ist, und der Algorithmus kann terminieren. Für $\Delta x^K \neq 0$ können wir hingegen den folgenden ausschlaggebenden Satz beweisen:

Satz 4.2.3. *Es sei ein quadratisches Optimierungsproblem $(\widetilde{\text{NLP}}^K)$ mit symmetrischer und positiv definiten Matrix L_K gegeben, wobei x^K die Voraussetzungen $Ax^K \geq \alpha$ und $Bx^K = \beta$ erfülle. Δx^K sei ein Minimum von $(\widetilde{\text{NLP}}^K)$ und es gelte $\Delta x^K \neq 0$. Dann ist*

$$\nabla f(x^K)^\top \Delta x^K \leq -(\Delta x^K)^\top L_K \Delta x^K < 0.$$

Beweis. (Siehe Geiger und Kanzow, 2002.)

Wir verwenden wieder die Notation $A = (a_1, \dots, a_p)^\top$ mit $a_i \in \mathbb{R}^n, \forall i = 1, \dots, p$ aus Kapitel 4.1.2. Damit lässt sich Bedingung $[\widetilde{\text{NLP}}^K\text{-iii}]$ auch als

$$\lambda_i^{K+1} \geq 0, \quad \lambda_i^{K+1} (a_i^\top x^K - \alpha_i + a_i^\top \Delta x^K) = 0, \quad \forall i = 1, \dots, p \quad (4.28)$$

schreiben. Wegen der Annahmen bezüglich x^K folgt aus $[\widetilde{\text{NLP}}^K\text{-i}]$, dass $B\Delta x^K = 0$ ist. Multiplizieren wir nun $[\widetilde{\text{NLP}}^K\text{-ii}]$ von links mit $(\Delta x^K)^\top$, so folgt

$$(\Delta x^K)^\top L_K \Delta x^K + \nabla f(x^K)^\top \Delta x^K - (\lambda^{K+1})^\top A \Delta x^K = 0. \quad (4.29)$$

Da für x^K nach Voraussetzung $a_i^\top x^K \geq \alpha_i$ gilt, können wir aus (4.28)

$$-(\lambda^{K+1})^\top A \Delta x^K = - \sum_{i=1}^p \lambda_i^{K+1} a_i^\top \Delta x^K = \sum_{i=1}^p \lambda_i^{K+1} (a_i^\top x^K - \alpha_i) \geq 0$$

folgern. Setzen wir dies in (4.29) ein, so erhalten wir wegen $L_K \succ 0$ und $\Delta x^K \neq 0$ insgesamt

$$\nabla f(x^K)^\top \Delta x^K \leq -(\Delta x^K)^\top L_K \Delta x^K < 0.$$

□

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

Damit ist Δx^K nach Lemma A.5.1 eine Abstiegsrichtung der Funktion f in x^K .

Unser globalisiertes SQP-Verfahren erzeugt nun Iterationspunkte $x^{K+1} := x^K + t^K \Delta x^K$ mit Schrittweiten $t^K \in (0, 1]$, wobei die in Appendix A.5 beschriebene Armijo-Regel zur Anwendung kommt. Dabei werden $\delta \in (0, 1)$ und $\tau \in (0, 1)$ vorgegeben und $t^K := \max\{\tau^l \mid l = 0, 1, \dots\}$ wird so ermittelt, dass

$$f(x^K + t^K \Delta x^K) \leq f(x^K) + \delta t^K \nabla f(x^K)^\top \Delta x^K. \quad (4.30)$$

Wegen Satz A.5.2 wissen wir, dass

$$\operatorname{argmax}_{l=0,1,\dots} \{\tau^l \mid (4.30) \text{ ist erfüllt}\} < \infty.$$

Daher ist auch $t^K > 0$ und $f(x^{K+1}) < f(x^K)$.

Starten wir ein solches Verfahren mit einem beliebigen, jedoch für $(\widetilde{\text{NLP}})$ zulässigen Punkt, so können wir induktiv zeigen, dass alle vom Algorithmus erzeugten Punkte zulässig bleiben:

- *Gleichheitsrestriktionen:*

Mit $Bx^K = \beta$ erfüllt eine Lösung Δx^K von $(\widetilde{\text{NLP}}^K)$ $B\Delta x^K = 0$. Damit ist $Bx^{K+1} = Bx^K + t^K B\Delta x^K = \beta$.

- *Ungleichheitsrestriktionen:*

Eine Lösung Δx^K von $(\widetilde{\text{NLP}}^K)$ erfüllt $A\Delta x^K \geq \alpha - Ax^K$. Es folgt mit der Annahme $Ax^K \geq \alpha$, dass

$$Ax^{K+1} = Ax^K + t^K A\Delta x^K \geq (1 - t^K)Ax^K + t^K \alpha \geq (1 - t^K)\alpha + t^K \alpha = \alpha.$$

Damit ist auch die Annahme aus Satz 4.2.3, dass x^K für $(\widetilde{\text{NLP}})$ zulässig ist, kein Problem mehr.

Wir haben zuvor festgestellt, dass wir im Fall $\Delta x^K = 0$ bereits einen KKT-Punkt von $(\widetilde{\text{NLP}})$ vorliegen haben. Umgekehrt wissen wir nun jedoch auch, dass im Fall $\Delta x^K \neq 0$ ein zulässiger Punkt mit kleinerem Zielfunktionswert existiert.

Wir fassen den resultierenden Algorithmus zusammen:

Algorithmus 4.7 *Ein zulässiges globales SQP-Verfahren*

Input:

- Eine Instanz des nichtlinearen Optimierungsproblems $(\widetilde{\text{NLP}})$, welche durch $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$, $A \in \mathbb{R}^{p \times n}$, $\alpha \in \mathbb{R}^p$, $B \in \mathbb{R}^{q \times n}$ und $\beta \in \mathbb{R}^q$ beschrieben wird.
- Ein Punkt $x^0 \in \mathbb{R}^n$, welcher zulässig für $(\widetilde{\text{NLP}})$ ist.
- Eine Berechnungsvorschrift zur Ermittlung symmetrischer und positiv definiten Matrizen $L_K \in \mathbb{R}^{n \times n}$, $K \in \{0, 1, \dots\}$.
- $\delta \in (0, 1)$ und $\tau \in (0, 1)$.

Output:

- Ein KKT-Punkt (x^*, λ, γ) von $(\widetilde{\text{NLP}})$.
-

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

- 1: Setze $K := 0$.
- 2: Ermittle $L_K \in \mathbb{R}^{n \times n}$.
- 3: Berechne einen KKT-Punkt $(\Delta x^K, \lambda^{K+1}, \gamma^{K+1})$ des quadratischen Programms $(\widetilde{\text{NLP}}^K)$.
- 4: **if** $\Delta x^K = 0$ **then**
- 5: Setze $x^{K+1} := x^K$.
- 6: **return** $(x^{K+1}, \lambda^{K+1}, \gamma^{K+1})$
- 7: **end if**
- 8: Berechne $t^K := \max\{\tau^l \mid l = 0, 1, \dots\}$, sodass

$$f(x^K + t^K \Delta x^K) \leq f(x^K) + \delta t^K \nabla f(x^K)^\top \Delta x^K.$$

- 9: Setze $x^{K+1} := x^K + t^K \Delta x^K$.
- 10: Setze $K := K + 1$.
- 11: **goto** Zeile 2.

Für diesen Algorithmus ist ausschlaggebend, dass es uns möglich ist, in jeder Iteration eine Abstiegsrichtung von f zu bestimmen. Damit ist auch gewährleistet, dass das Verfahren nicht in einem KKT-Punkt von $(\widetilde{\text{NLP}})$ landet, der einem *Maximum* der Zielfunktion auf der zulässigen Menge entspricht.

Für Optimierungsprobleme mit nichtlinearen Restriktionen können durch obige Vorgangsweise im Allgemeinen auch unzulässige Iterationspunkte auftreten und Satz 4.2.3 ist in dieser Form nicht anwendbar. Stattdessen werden Abstiegsrichtungen einer sogenannten Penalty-Funktion berechnet, wobei es möglich ist, für die *exakte* ℓ_1 -Penalty-Funktion ein analoges Resultat zu Satz 4.2.3 zu beweisen (siehe Geiger und Kanzow, 2002). Diese Penalty-Funktion stimmt für Probleme mit linearen Restriktionen und zulässigen Punkten mit f überein.

Es wäre nun wünschenswert, wenn die globale SQP-Methode (in Analogie zu unserem lokalen SQP-Verfahren) ebenfalls lokal mit superlinearer oder quadratischer Konvergenzrate konvergieren würde. Verwenden wir beispielsweise für L_K die Hesse-Matrix der Lagrangefunktion von $(\widetilde{\text{NLP}})$, welche aufgrund der linearen Restriktionen gleich $\nabla^2 f(x^K)$ ist, und nehmen an, dass diese Matrix lokal positiv semidefinit ist, womit in jeder Iteration der berechnete KKT-Punkt des quadratischen Teilproblems *eindeutig* ist, so können wir aus Satz 4.2.2 (ohne Berücksichtigung der dortigen Norm-Einschränkung) für die globale SQP-Methode lokale Konvergenz mit der entsprechenden Konvergenzrate folgern, falls *lokal* die *volle* Schrittweite $t^K = 1$ akzeptiert wird. Für unseren Fall, in dem nur lineare Restriktionen vorliegen und jede beliebige Schrittweite $t^K \leq 1$ stets nur zulässige Iterationspunkte erzeugt, gilt nach Sun und Yuan (2006), dass lokal $f(x^K + \Delta x^K) < f(x^K)$ ist und ein voller Schritt mit $t^K = 1$ durchgeführt werden kann. Es sei erwähnt, dass dies für allgemeine nichtlineare Probleme (bei Verwendung der bereits erwähnten Penalty-Funktion) nicht garantiert ist. Geiger und Kanzow (2002) sowie Nocedal und Wright (2006) geben hierfür Beispiele mit nichtlinearen Restriktionen an. Dieses Verhalten wird nach Maratos (1978) als

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

„Maratos-Effekt“ bezeichnet. Möglichkeiten zur Vermeidung dieses Effektes werden unter anderem von Sun und Yuan (2006) beschrieben.

Wir zeigen nun noch ein **globales Konvergenzresultat** für unser globales SQP-Verfahren. Als Referenz dienen uns wieder Geiger und Kanzow (2002), die das folgende Resultat für eine *modifizierte* Version des SQP-Verfahrens angeben. Die Verwendung dieser Modifikation führt einerseits dazu, dass die quadratischen Teilprobleme (NLP^K) für beliebige nichtlineare Programme (NLP) stets lösbar sind, und verhindert andererseits ein unbeschränktes Anwachsen der Lagrange-Multiplikatoren im Verlauf des Algorithmus, was von großer Bedeutung ist. Wir übertragen die Ergebnisse von Geiger und Kanzow (2002) hingegen wieder auf unseren globalen SQP-Algorithmus und setzen die Beschränktheit der Lagrange-Multiplikatoren vorerst voraus. Bevor wir jedoch den entsprechenden Konvergenzsatz angeben können, benötigen wir noch nachstehende Definition:

Definition. (Gleichmäßig positiv definite und beschränkte Matrizen, vgl. Alt (2010))
Eine Folge $\{L_K\}_{K=0,1,\dots}$ von symmetrischen Matrizen mit $L_K \in \mathbb{R}^{n \times n}$ heißt *gleichmäßig positiv definit und beschränkt*, falls für alle $K = 0, 1, \dots$ und für alle $x \in \mathbb{R}^n$

$$c_1 \|x\| \leq x^\top L_K x \leq c_2 \|x\|$$

für zwei Konstanten $c_2 \geq c_1 > 0$ gilt.

Nach Geiger und Kanzow (1999) gilt im Zusammenhang mit dieser Klasse von Matrizen folgende Eigenschaft:

Lemma 4.2.4. *Für eine beliebige Folge $\{L_K\}_{K=0,1,\dots}$ von symmetrischen und positiv definiten Matrizen sind die folgenden drei Aussagen äquivalent:*

1. Die Folgen $\{L_K\}_{K=0,1,\dots}$ sowie $\{L_K^{-1}\}_{K=0,1,\dots}$ sind beschränkt.
2. Die Folge $\{L_K\}_{K=0,1,\dots}$ ist gleichmäßig positiv definit und beschränkt.
3. Die Folge $\{L_K^{-1}\}_{K=0,1,\dots}$ ist gleichmäßig positiv definit und beschränkt.

Beweis. Siehe Geiger und Kanzow (1999). □

Satz 4.2.5. *(Konvergenz des zulässigen globalen SQP-Verfahrens)*

Die Zielfunktion f von $(\widetilde{\text{NLP}})$ sei auf \mathbb{R}^n stetig differenzierbar. Wir betrachten den zulässigen globalen SQP-Algorithmus 4.7 für eine gleichmäßig positiv definite und beschränkte Folge von symmetrischen Matrizen $\{L_K\}_{K=0,1,\dots}$ und nehmen an, dass für die in der 3. Zeile des Algorithmus berechneten Punkte $(\Delta x^K, \lambda^{K+1}, \gamma^{K+1})$ die Folgen $\{\Delta x^K\}_{K=0,1,\dots}$, $\{\lambda^{K+1}\}_{K=0,1,\dots}$ und $\{\gamma^{K+1}\}_{K=0,1,\dots}$ beschränkt sind. Dann existieren für jeden Häufungspunkt \hat{x} der Folge $\{x^K\}_{K=0,1,\dots}$ von erzeugten Iterationspunkten Lagrange-Multiplikatoren $\hat{\lambda}$ und $\hat{\gamma}$, sodass $(\hat{x}, \hat{\lambda}, \hat{\gamma})$ ein KKT-Punkt von $(\widetilde{\text{NLP}})$ ist.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

Beweis. (Nach Geiger und Kanzow, 2002.)

Sei \hat{x} ein Häufungspunkt der Folge $\{x^K\}_{K=0,1,\dots}$. Dann existiert eine Teilfolge $\{x^{K_i}\}_{i=0,1,\dots}$, die gegen \hat{x} konvergiert. Laut Annahme sind die Lagrange-Multiplikatoren beschränkt. Daher existiert eine Teilfolge $\{\lambda^{K_{i_j}+1}\}_{j=0,1,\dots}$ von $\{\lambda^{K_i+1}\}_{i=0,1,\dots}$, die gegen einen Punkt $\hat{\lambda}$ konvergiert. Weiters existiert eine Teilfolge $\{\gamma^{K_{i_j k}+1}\}_{k=0,1,\dots}$ von $\{\gamma^{K_{i_j}+1}\}_{j=0,1,\dots}$, die gegen einen Punkt $\hat{\gamma}$ konvergiert. Nach Lemma 4.2.4 sind sowohl $\{L_K\}_{K=0,1,\dots}$ als auch $\{L_K^{-1}\}_{K=0,1,\dots}$ beschränkt. Es gibt also wiederum eine Teilfolge $\{L_{K_{i_j k_m}}\}_{m=0,1,\dots}$ von $\{L_{K_{i_j k}}\}_{k=0,1,\dots}$, die gegen eine symmetrische und positiv definite Matrix \hat{L} konvergiert. Darüber hinaus ist laut Annahme die Folge $\{\Delta x^K\}_{K=0,1,\dots}$ und somit auch $\{\|\Delta x^{K_{i_j k_m}}\|\}_{m=0,1,\dots}$ beschränkt. $\{\|\Delta x^{K_{i_j k_m l}}\|\}_{l=0,1,\dots}$ sei nun diejenige Teilfolge, die gegen den *kleinsten* Häufungspunkt von $\{\|\Delta x^{K_{i_j k_m}}\|\}_{m=0,1,\dots}$ konvergiert. Diese Wahl ist legitim, da ein kleinster Häufungspunkt aufgrund der Abgeschlossenheit der Menge aller Häufungspunkte existiert.

Wir setzen nun zur Notationsvereinfachung $\hat{K}_l := K_{i_j k_m l}$, $\forall l \in \{0, 1, \dots\}$, und arbeiten mit dieser Indexfolge, für die laut Konstruktion

$$\lim_{l \rightarrow \infty} x^{\hat{K}_l} = \hat{x}, \quad \lim_{l \rightarrow \infty} \lambda^{\hat{K}_l+1} = \hat{\lambda}, \quad \lim_{l \rightarrow \infty} \gamma^{\hat{K}_l+1} = \hat{\gamma}, \quad \lim_{l \rightarrow \infty} L_{\hat{K}_l} = \hat{L}, \quad (4.31)$$

gilt, weiter.

Wegen Zeile 8 in Algorithmus 4.7 ist die Folge $\{f(x^K)\}_{K=0,1,\dots}$ monoton fallend. Darüber hinaus konvergiert die Teilfolge $\{f(x^{\hat{K}_l})\}_{l=0,1,\dots}$ gegen $f(\hat{x})$, womit (aufgrund der Monotonie) auch die Folge $\{f(x^K)\}_{K=0,1,\dots}$ gegen $f(\hat{x})$ konvergiert und

$$\lim_{K \rightarrow \infty} (f(x^{K+1}) - f(x^K)) = 0$$

gilt. Weiters ist

$$f(x^{K+1}) - f(x^K) \leq \delta t^K \nabla f(x^K)^\top \Delta x^K \stackrel{\text{Satz}}{\leq} -\delta t^K (\Delta x^K)^\top L_K \Delta x^K < 0$$

und damit wegen $\delta \in (0, 1)$

$$\lim_{K \rightarrow \infty} t^K (\Delta x^K)^\top L_K \Delta x^K = 0. \quad (4.32)$$

Wir zeigen nun nachstehende Aussage:

Behauptung: Es gilt

$$\lim_{l \rightarrow \infty} \|\Delta x^{\hat{K}_l}\| = 0,$$

womit der kleinste Häufungspunkt der Folge $\{\|\Delta x^K\|\}_{K=0,1,\dots}$ gleich 0 ist und

$$\lim_{l \rightarrow \infty} \Delta x^{\hat{K}_l} = 0$$

folgt.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

Beweis: Wir beweisen die Aussage durch Widerspruch, indem wir annehmen, dass

$$\lim_{l \rightarrow \infty} \|\Delta x^{\hat{K}_l}\| =: s > 0.$$

Damit kann jedoch (4.32) nur dann erfüllt sein, wenn $\lim_{l \rightarrow \infty} t^{\hat{K}_l} = 0$. Dies bedeutet, dass ein $\tilde{l} \in \{0, 1, \dots\}$ existiert, sodass $t^{\hat{K}_l} < 1$ ist für alle $l \geq \tilde{l}$. Betrachten wir Zeile 8 des globalen SQP-Algorithmus, so ist ersichtlich, dass für alle $l \geq \tilde{l}$

$$f\left(x^{\hat{K}_l} + \frac{t^{\hat{K}_l}}{\tau} \Delta x^{\hat{K}_l}\right) > f(x^{\hat{K}_l}) + \delta \frac{t^{\hat{K}_l}}{\tau} \nabla f(x^{\hat{K}_l})^\top \Delta x^{\hat{K}_l} \quad (4.33)$$

gilt. Wir verwenden nun die Notation $\bar{t}^{\hat{K}_l} := \frac{t^{\hat{K}_l}}{\tau}$. Unter Anwendung des Mittelwertsatzes der Differentialrechnung folgt mit $\xi^{\hat{K}_l} \in (0, 1)$, dass

$$\begin{aligned} f(x^{\hat{K}_l} + \bar{t}^{\hat{K}_l} \Delta x^{\hat{K}_l}) &= f(x^{\hat{K}_l}) + \bar{t}^{\hat{K}_l} \nabla f(x^{\hat{K}_l} + \xi^{\hat{K}_l} \bar{t}^{\hat{K}_l} \Delta x^{\hat{K}_l})^\top \Delta x^{\hat{K}_l} \\ &= f(x^{\hat{K}_l}) + \bar{t}^{\hat{K}_l} \nabla f(x^{\hat{K}_l})^\top \Delta x^{\hat{K}_l} \\ &\quad + \bar{t}^{\hat{K}_l} (\nabla f(x^{\hat{K}_l} + \xi^{\hat{K}_l} \bar{t}^{\hat{K}_l} \Delta x^{\hat{K}_l}) - \nabla f(x^{\hat{K}_l}))^\top \Delta x^{\hat{K}_l} \\ &= f(x^{\hat{K}_l}) + \bar{t}^{\hat{K}_l} \nabla f(x^{\hat{K}_l})^\top \Delta x^{\hat{K}_l} + r^{\hat{K}_l}, \end{aligned}$$

wobei $r^{\hat{K}_l} := \bar{t}^{\hat{K}_l} (\nabla f(x^{\hat{K}_l} + \xi^{\hat{K}_l} \bar{t}^{\hat{K}_l} \Delta x^{\hat{K}_l}) - \nabla f(x^{\hat{K}_l}))^\top \Delta x^{\hat{K}_l}$ bezeichnet. Es gilt nun $\lim_{l \rightarrow \infty} \frac{r^{\hat{K}_l}}{\bar{t}^{\hat{K}_l}} = 0$, da f stetig differenzierbar ist und $\lim_{l \rightarrow \infty} \bar{t}^{\hat{K}_l} = \frac{1}{\tau} \lim_{l \rightarrow \infty} t^{\hat{K}_l} = 0$ ist. Setzen wir dieses Ergebnis in (4.33) ein und berücksichtigen, dass für eine Konstante $c > 0$ die Ungleichung $x^\top L_K x \geq c\|x\|$ erfüllt ist, so erhalten wir mit $\delta \in (0, 1)$ folgendes Resultat:

$$\begin{aligned} \bar{t}^{\hat{K}_l} \nabla f(x^{\hat{K}_l})^\top \Delta x^{\hat{K}_l} + r^{\hat{K}_l} &> \delta \bar{t}^{\hat{K}_l} \nabla f(x^{\hat{K}_l})^\top \Delta x^{\hat{K}_l} \\ &\Leftrightarrow \\ \frac{r^{\hat{K}_l}}{\bar{t}^{\hat{K}_l}} &> (\delta - 1) \nabla f(x^{\hat{K}_l})^\top \Delta x^{\hat{K}_l} \stackrel{\text{Satz 4.2.3}}{\geq} (1 - \delta) (\Delta x^{\hat{K}_l})^\top L_K \Delta x^{\hat{K}_l} \geq c(1 - \delta) \|\Delta x^{\hat{K}_l}\| \end{aligned}$$

Mit $l \rightarrow \infty$ gilt schlussendlich

$$c(1 - \delta)s \leq 0.$$

Da jedoch $c(1 - \delta) > 0$ gilt, ist dies ein Widerspruch zu $s > 0$, womit

$$\lim_{l \rightarrow \infty} \|\Delta x^{\hat{K}_l}\| = 0$$

folgt.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

Mit obiger Behauptung konvergiert also $\{\Delta x^{\hat{K}_l}\}_{l=0,1,\dots}$ gegen 0. Unter Verwendung dieser Aussage und (4.31) erhalten wir nun aus den KKT-Bedingungen $[\widetilde{\text{NLP}}^K\text{-i}]$, $[\widetilde{\text{NLP}}^K\text{-ii}]$ und $[\widetilde{\text{NLP}}^K\text{-iii}]$ auf der Teilfolge $\{\hat{K}_l\}_{l=0,1,\dots}$ mit $l \rightarrow \infty$ das nachstehende Gleichungssystem:

- $A\hat{x} \geq \alpha$, $B\hat{x} = \beta$,
- $\nabla f(\hat{x}) - A^\top \hat{\lambda} - B^\top \hat{\gamma} = 0$,
- $\hat{\lambda} \geq 0$, $\hat{\lambda}^\top \cdot (A\hat{x} - \alpha) = 0$.

Aus diesem Grund ist $(\hat{x}, \hat{\lambda}, \hat{\gamma})$ ein KKT-Punkt des ursprünglichen Problems $(\widetilde{\text{NLP}})$. \square

Alt (2010) gibt eine Bedingung an, unter der die Beschränktheit der Lagrange-Multiplikatoren für allgemeine nichtlineare Probleme erfüllt ist. Wir formulieren dieses Kriterium nun für unsere eingeschränkte Problemklasse mit linearen Restriktionen, wobei wir die Notationen (4.2) und (4.3) aus Kapitel 4.1.2 verwenden:

Lemma 4.2.6. *Für eine symmetrische Matrix $L \in \mathbb{R}^{n \times n}$ mit $L \succ 0$ und eine Matrix $B \in \mathbb{R}^{q \times n}$ mit $\text{rank}(B) = q$ ist die Matrix*

$$M := \begin{pmatrix} L & B^\top \\ B & 0 \end{pmatrix}$$

invertierbar. Es gilt

$$M^{-1} = \begin{pmatrix} L^{-1} - L^{-1}B^\top H^{-1}BL^{-1} & L^{-1}B^\top H^{-1} \\ H^{-1}BL^{-1} & -H^{-1} \end{pmatrix}$$

mit $H = BL^{-1}B^\top$.

Beweis. Siehe Alt (2010). \square

Satz 4.2.7. *Die Zielfunktion f von $(\widetilde{\text{NLP}})$ sei auf \mathbb{R}^n stetig differenzierbar. Wir betrachten den zulässigen globalen SQP-Algorithmus 4.7 für eine Folge von symmetrischen Matrizen $\{L_K\}_{K=0,1,\dots}$, die gleichmäßig positiv definit und beschränkt ist. Für den (zulässigen) Startpunkt $x^0 \in \mathbb{R}^n$ sei die Menge $N_{\widetilde{\text{NLP}}}(x^0) := \{x \in \mathbb{R}^n \mid Ax \geq \alpha, Bx = \beta \text{ und } f(x) \leq f(x^0)\}$ kompakt.²¹ Ist nun in der $(K+1)$ -ten Iteration Δx^K die (eindeutige) Lösung des quadratischen Problems $(\widetilde{\text{NLP}}^K)$ und sind die Vektoren $a_j, \forall j \in J(\Delta x^K)$ und $b_k, \forall k \in \{1, \dots, q\}$ linear unabhängig,²² so sind auch die Lagrange-Multiplikatoren λ^{K+1} und γ^{K+1} eindeutig bestimmt und sowohl $\{\Delta x^K\}_{K=0,1,\dots}$ als auch $\{\lambda^{K+1}\}_{K=0,1,\dots}$ und $\{\gamma^{K+1}\}_{K=0,1,\dots}$ sind beschränkt.*

²¹Alt (2010) nennt $N_{\widetilde{\text{NLP}}}(x^0)$ auch *Niveaumenge*.

²² $J(\Delta x^K) = \{i \in \{1, \dots, p\} \mid a_i^\top \Delta x^K = \alpha_i - a_i^\top x^K\}$ bezeichnet wieder die Menge der aktiven Restriktionen.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

Beweis. (Vgl. Alt, 2010.)

Es gilt $\lambda_j^{K+1} = 0$ für alle $j \in J(\Delta x^K)^c := \{1, \dots, p\} \setminus J(\Delta x^K)$, womit diese Lagrange-Multiplikatoren eindeutig bestimmt sind. Der Vektor $(\Delta x^K, \lambda_{J(\Delta x^K)}^{K+1}, \gamma^{K+1})$ ist hingegen Lösung des linearen Gleichungssystems

$$\begin{pmatrix} L_K & -A_{J(\Delta x^K)}^\top & -B^\top \\ -A_{J(\Delta x^K)} & 0 & 0 \\ -B & 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \lambda \\ \gamma \end{pmatrix} = \begin{pmatrix} -\nabla f(x^K) \\ A_{J(\Delta x^K)} x^K - \alpha_{J(\Delta x^K)} \\ 0 \end{pmatrix}. \quad (4.34)$$

Dieses System ergibt sich aus den KKT-Bedingungen, wobei wir die Restriktionsbedingungen so umgeschrieben haben, dass auf der linken Seite eine symmetrische Matrix vorliegt. Wir bezeichnen nun die Koeffizientenmatrix mit M_K und den Vektor der rechten Seite mit m_K . Die Invertierbarkeit von M_K und damit die eindeutige Lösbarkeit von (4.34) ergibt sich aus Lemma 4.2.6.²³ Da die Folge $\{L_K^{-1}\}_{K=0,1,\dots}$ nach Lemma 4.2.4 gleichmäßig positiv und beschränkt ist, können wir aus Lemma 4.2.6 auch die Beschränktheit der Matrizenfolge $\{M_K^{-1}\}_{K=0,1,\dots}$ (d.h. $\|M_K^{-1}\| \leq c, \forall K = 0, 1, \dots$ mit $c > 0$) folgern. Damit erhalten wir

$$\left\| \begin{pmatrix} \Delta x^K \\ \lambda_{J(\Delta x^K)}^{K+1} \\ \gamma^{K+1} \end{pmatrix} \right\| = \|M_K^{-1} m_K\| \leq \|M_K^{-1}\| \cdot \|m_K\| \leq c \cdot \|m_K\|.$$

Wie wir bereits gesehen haben, sind sämtliche Iterationspunkte von Algorithmus 4.7 zulässig und liegen darüber hinaus in der kompakten Menge $N_{\widetilde{\text{NLP}}}(x^0)$, da im Laufe des Algorithmus keine Vergrößerung des Zielfunktionswertes erfolgen kann. Daraus folgt wegen der *stetigen* Differenzierbarkeit der Funktion f die Beschränktheit von m_K und somit auch von $\Delta x^K, \lambda^{K+1}$ und γ^{K+1} . \square

4.2.4. Eine modifizierte BFGS-Formel

Abschließend betrachten wir noch eine Möglichkeit, die Matrizen L_K als Approximation der Hesse-Matrix der Lagrangefunktion

$$\widetilde{\mathcal{L}}(x, \lambda, \gamma) := f(x) - \lambda^\top (Ax - \alpha) - \gamma^\top (Bx - \beta)$$

von $(\widetilde{\text{NLP}})$ so festzulegen, dass (ausgehend von $L_0 \succ 0$) ebendiese stets positiv definit bleiben. Eine Möglichkeit, dies zu realisieren, basiert auf der BFGS-Formel (**B**royden-**F**letcher-**G**oldfarb-**S**hanno-Formel), deren Herleitung im Rahmen der Quasi-Newton-Verfahren in Appendix A.7 zu finden ist. Wir definieren dazu die folgenden beiden Vektoren:

$$\begin{aligned} v^K &:= x^{K+1} - x^K = t^K \Delta x^K \\ w^K &:= \nabla_x \widetilde{\mathcal{L}}(x, \lambda, \gamma)|_{x=x^{K+1}} - \nabla_x \widetilde{\mathcal{L}}(x, \lambda, \gamma)|_{x=x^K} = \nabla f(x^{K+1}) - \nabla f(x^K) \end{aligned}$$

²³Alternativ lässt sich dieser Teil des Beweises analog zu Satz 4.1.2 a) in Kapitel 4.1.2 zeigen.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

w^K hängt dabei (ebenso wie die zu approximierende Matrix $\nabla_{xx}^2 \widetilde{\mathcal{L}}(x, \lambda, \gamma)|_{x=x^K} = \nabla^2 f(x^K)$) nicht von den Lagrange-Multiplikatoren λ und γ ab, was auf die Linearität der Restriktionen zurückzuführen ist. Daher könnten wir nun die Matrizenfolge $\{L_K^{-1}\}_{K=0,1,\dots}$ durch die BFGS-Formel aus Satz A.7.3 ermitteln.

Eine entscheidende Voraussetzung dafür, dass die resultierenden Matrizen auch wirklich positiv definit sind, ist jedoch $(v^K)^\top w^K > 0$ für alle $K \geq 1$. Dies kann für spezielle Schrittweitenregelungen zwar auch erreicht werden,²⁴ nicht jedoch für jene Version der Armijo-Regel, wie wir sie benutzen. Wir halten uns daher an Geiger und Kanzow (2002) und verwenden die folgende, von Powell (1978) vorgeschlagene Modifikation der BFGS-Formel:

Sei

$$z^K := \vartheta^K w^K + (1 - \vartheta^K) L_K v^K \quad (4.35)$$

mit

$$\vartheta^K := \begin{cases} 1, & \text{falls } (v^K)^\top w^K \geq 0.2(v^K)^\top L_K v^K, \\ \frac{0.8(v^K)^\top L_K v^K}{(v^K)^\top L_K v^K - (v^K)^\top w^K}, & \text{falls } (v^K)^\top w^K < 0.2(v^K)^\top L_K v^K. \end{cases} \quad (4.36)$$

Dann ist die modifizierte BFGS-Formel nach Powell durch

$$L_{K+1} := L_K + \frac{z^K (z^K)^\top}{(v^K)^\top z^K} - \frac{L_K v^K (v^K)^\top L_K}{(v^K)^\top L_K v^K} \quad (4.37)$$

gegeben. Hier ersetzt also der Vektor z^K den Vektor w^K in der ursprünglichen BFGS-Formel (A.6).

Es lässt sich nun nachstehende Aussage formulieren:

Lemma 4.2.8. *$L_K \in \mathbb{R}^{n \times n}$ sei eine symmetrische und positiv definite Matrix. Für zwei Vektoren $v^K \in \mathbb{R}^n \setminus \{0\}$ und $w^K \in \mathbb{R}^n$ sei der Vektor $z^K \in \mathbb{R}^n$ durch (4.35) und (4.36) definiert. Dann ist*

$$(v^K)^\top z^K > 0$$

und die durch (4.37) definierte Matrix $L_{K+1} \in \mathbb{R}^{n \times n}$ ist symmetrisch und positiv definit.

Beweis.

- Für den Beweis von $(v^K)^\top z^K > 0$ unterscheiden wir zwei Fälle (vgl. Geiger und Kanzow, 2002):

$$- (v^K)^\top w^K \geq 0.2(v^K)^\top L_K v^K:$$

Dann ist $\vartheta^K = 1$ und $z^K = w^K$. Es folgt

$$(v^K)^\top z^K = (v^K)^\top w^K \geq 0.2(v^K)^\top L_K v^K > 0,$$

da L_K laut Annahme positiv definit ist und $v^K \neq 0$.

²⁴Ein Beweis dieser Aussage ist für die *Wolfe-Powell-Schrittweitenregelung* in Geiger und Kanzow (1999) angegeben.

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

$$- (v^K)^\top w^K < 0.2(v^K)^\top L_K v^K:$$

Per Definition von z^K gilt

$$\begin{aligned} (v^K)^\top z^K &= \frac{0.8(v^K)^\top L_K v^K}{(v^K)^\top L_K v^K - (v^K)^\top w^K} (v^K)^\top w^K \\ &\quad + \frac{0.2(v^K)^\top L_K v^K - (v^K)^\top w^K}{(v^K)^\top L_K v^K - (v^K)^\top w^K} (v^K)^\top L_K v^K \\ &= \frac{0.8(v^K)^\top w^K + 0.2(v^K)^\top L_K v^K - (v^K)^\top w^K}{(v^K)^\top L_K v^K - (v^K)^\top w^K} (v^K)^\top L_K v^K \\ &= 0.2(v^K)^\top L_K v^K \\ &> 0. \end{aligned}$$

- Wir zeigen nun, dass die (offensichtlich symmetrische) Matrix L_{K+1} positiv definit ist (siehe Geiger und Kanzow, 1999):

Wegen Lemma A.7.2 existiert eine symmetrische und positiv definite Matrix $L_K^{1/2} \in \mathbb{R}^{n \times n}$ mit $L_K^{1/2} L_K^{1/2} = L_K$. Verwenden wir nun die Definition (4.37) von L_{K+1} , so erhalten wir für ein $x \in \mathbb{R}^n \setminus \{0\}$ mit der Notation $\tilde{v}^K := L_K^{1/2} v^K$ und $\tilde{x}^K := L_K^{1/2} x$ die Aussage

$$\begin{aligned} x^\top L_{K+1} x &= x^\top L_K x + \frac{(x^\top z^K)^2}{(v^K)^\top z^K} - \frac{(x^\top L_K v^K)^2}{(v^K)^\top L_K v^K} \\ &= (\tilde{x}^K)^\top \tilde{x}^K + \frac{(x^\top z^K)^2}{(v^K)^\top z^K} - \frac{((\tilde{x}^K)^\top \tilde{v}^K)^2}{(\tilde{v}^K)^\top \tilde{v}^K} \\ &\stackrel{(1)}{\geq} \|\tilde{x}^K\|^2 + \frac{(x^\top z^K)^2}{(v^K)^\top z^K} - \frac{\|\tilde{x}^K\|^2 \|\tilde{v}^K\|^2}{\|\tilde{v}^K\|^2} = \frac{(x^\top z^K)^2}{(v^K)^\top z^K} \\ &\stackrel{(2)}{\geq} 0. \end{aligned}$$

Ungleichung (1) folgt dabei aus der Cauchy-Schwarz'schen Ungleichung (siehe Appendix A.4), welche besagt, dass $((\tilde{x}^K)^\top \tilde{v}^K)^2 \leq \|\tilde{x}^K\|^2 \|\tilde{v}^K\|^2$, und Ungleichung (2) ergibt sich unmittelbar aus dem ersten Teil des Beweises. L_{K+1} ist also positiv semidefinit und kann nur dann *nicht* positiv definit sein, wenn die beiden Ungleichungen (1) und (2) für mindestens ein $x \neq 0$ mit Gleichheit erfüllt sind.

Wir nehmen nun also an, dass (für $x \neq 0$) in Ungleichung (1) Gleichheit herrscht, womit nach Satz A.4.1 \tilde{x}^K und \tilde{v}^K linear abhängig sind. Unter Berücksichtigung der Definitionen von \tilde{x}^K und \tilde{v}^K und der Tatsache, dass $L_K^{1/2}$ positiv definit ist, können wir folgern, dass auch x und v^K linear abhängig sind.

Unter der Annahme, dass nun auch Ungleichung (2) mit Gleichheit erfüllt ist, muss $x^\top z^K$ gleich 0 sein. Damit ist jedoch aufgrund der linearen Abhängigkeit von x und v^K auch $(v^K)^\top z^K = 0$. Dies ist ein Widerspruch zur Aussage $(v^K)^\top z^K > 0$ aus dem ersten Teil des Beweises. Daher ist $L_{K+1} \succ 0$.

□

4. Algorithmen zur Lösung nichtlinearer Optimierungsprobleme

Wir können also durch die Wahl der Matrizen L_K mithilfe der modifizierten BFGS-Formel die wichtige Bedingung der positiven Definitheit sicherstellen. Laut Spellucci (1993) war allerdings im Erscheinungsjahr 1993 dessen Buches „*Numerische Verfahren der nichtlinearen Optimierung*“ nicht bekannt, ob die durch die modifizierte BFGS-Formel (4.37) definierte Folge $\{L_K\}_{K=0,1,\dots}$ auch gleichmäßig positiv definit und beschränkt ist. Uns ist es ebenfalls nicht gelungen, diese Eigenschaft zu zeigen oder zu widerlegen.

Geiger und Kanzow (2002) beschreiben die modifizierte BFGS-Formel als „Interpolation“ zwischen der Matrix L_K aus dem letzten Schritt und der durch die BFGS-Formel definierten Matrix in (A.6). Diese Interpretation ergibt sich aus folgenden Beobachtungen:

- Ist $(v^K)^\top w^K \geq 0.2(v^K)^\top L_K v^K$, womit in (4.36) der erste Fall eintritt, so ist $\vartheta^K = 1$ und es resultiert wegen $z^K = w^K$ die BFGS-Formel ohne Modifikation.
- Tritt hingegen in (4.36) der zweite Fall ein, so resultiert für $v^K \neq 0$ stets $\vartheta \in (0, 1)$, da bei der Berechnung von ϑ der Zähler $0.8(v^K)^\top L_K v^K$ aufgrund der positiven Definitheit von L_K stets positiv ist und der Nenner $(v^K)^\top L_K v^K - (v^K)^\top w^K$ wegen $(v^K)^\top w^K < 0.2(v^K)^\top L_K v^K$ stets größer als der Zähler ist. Ist $(v^K)^\top w^K$ so klein, dass wir $\vartheta^K \approx 0$ erhalten, folgt $z^K \approx L_K v^K$ und damit

$$L_{K+1} \approx L_K + \frac{L_K v^K (v^K)^\top L_K}{(v^K)^\top L_K v^K} - \frac{L_K v^K (v^K)^\top L_K}{(v^K)^\top L_K v^K} = L_K.$$

In diesem Fall wird also die vorherige Matrix L_K annähernd beibehalten.

5. Portfoliooptimierung in der Praxis

5.1. Durchgeführte Implementierungen

Sämtliche Implementierungen wurden in *GNU Octave*, Version 3.8.1, unter Verwendung des Betriebssystems Ubuntu 14.04 LTS durchgeführt. Es wurden einige *Octave-Forge*-Zusatzpakete installiert, welche gegebenenfalls für den korrekten Einsatz der in dieser Arbeit verwendeten Codes notwendig sind. Informationen dazu sind unter <http://octave.sourceforge.net/> zu finden. Die folgenden Pakete wurden eingebunden:

- `financial`, Version 0.4.0
- `general`, Version 1.3.4
- `io`, Version 2.2.6
- `miscellaneous`, Version 1.2.1
- `optim`, Version 1.4.0
- `statistics`, Version 1.2.3
- `struct`, Version 1.0.10
- `symbolic`, Version 1.1.0

Es wurden im Wesentlichen folgende Optimierungsmethoden implementiert:

- Die Methode der aktiven Menge aus Abschnitt 4.1.2
- Ein unzulässiges Innere-Punkte-Verfahren, und zwar das „Predictor-Corrector“-Verfahren nach Mehrotra (1992) aus Abschnitt 4.1.3
- Das unzulässige globale SQP-Verfahren aus Abschnitt 4.2.3

Dabei ist zu beachten, dass unsere Methode der aktiven Menge zur Lösung eines konvexen quadratischen Optimierungsproblems der Form (QP) dient, während Mehrotras Innere-Punkte-Verfahren auf der äquivalenten Formulierung ($\widetilde{\text{QP}}$) basiert.²⁵

Die Umsetzung der drei Verfahren orientiert sich an den bereits theoretisch untersuchten Pseudoalgorithmen und wird in einer separaten Projektarbeit genauer besprochen. Es seien an dieser Stelle nur einige Besonderheiten der konkreten Implementierung des globalen SQP-Verfahrens, welches die Behandlung von nichtlinearen Optimierungsproblemen ($\widetilde{\text{NLP}}$) mit linearen Gleichheits- und Ungleichheitsrestriktionen ermöglicht,²⁶ erwähnt:

²⁵(QP) und ($\widetilde{\text{QP}}$) sind in Abschnitt 1.3 nachzulesen.

²⁶($\widetilde{\text{NLP}}$) wird am Beginn von Kapitel 4.2 definiert.

5. Portfoliooptimierung in der Praxis

- Pseudoalgorithmus 4.7 setzt eigentlich einen zulässigen Startpunkt voraus. Es kann jedoch nicht nur mit zulässigen, sondern auch mit unzulässigen Startpunkten begonnen werden. Wird ein unzulässiger Punkt x^0 verwendet, so wird in der ersten Iteration trotzdem eine Lösung Δx^0 des quadratischen Programms $(\widetilde{\text{NLP}}^0)$ (Zeile 3 des Pseudoalgorithmus) ermittelt. Anschließend wird $x^1 := x^0 + \Delta x^0$ gesetzt. Dies bedeutet also, dass die Schrittweitenregelung in Zeile 8 übergangen wird. Dadurch wird offensichtlich erreicht, dass der neue Punkt x^1 zulässig ist, und die folgenden Iterationen können nach dem Schema des Pseudoalgorithmus fortgesetzt werden. Diese Implementierung bietet den Vorteil, einen eventuell bekannten, „guten“ (also in der Nähe der Optimallösung liegenden) Startpunkt verwenden zu können, auch wenn dieser vielleicht unzulässig ist.
- Es ist frei wählbar, ob die konvexen quadratischen Teilprobleme $(\widetilde{\text{NLP}}^K)$, die in jeder Iteration des Algorithmus auftreten (siehe Pseudoalgorithmus 4.7), mit der Methode der aktiven Menge oder Mehrotras „Predictor-Corrector“-Verfahren gelöst werden.
- Der Algorithmus ist in der Lage, für die Matrix L_K der Teilprobleme $(\widetilde{\text{NLP}}^K)$ immer dann die Hesse-Matrix $\nabla^2 f(x^K)$ zu verwenden, falls diese positiv definit ist. Ansonsten wird auf jene Matrix zurückgegriffen, welche durch die modifizierte BFGS-Formel (**B**royden-**F**letcher-**G**oldfarb-**S**hanno-Formel, siehe Abschnitt 4.2.4) gegeben ist. Die (symmetrische, positiv definite) Startmatrix für die BFGS-Formel ist defaultmäßig die n -dimensionale Einheitsmatrix, kann jedoch auch frei gewählt werden. Der Algorithmus liefert unter anderem auch die in der letzten Iteration berechnete BFGS-Matrix als Output, womit diese gegebenenfalls weiterverwendet werden kann.

Aufgrund der letzten Bemerkung bzw. der Tatsache, dass $(\widetilde{\text{NLP}}^K)$ den Gradienten $\nabla f(x^K)$ beinhaltet, interessieren wir uns dafür, wie die Gradienten und Hesse-Matrizen der für unsere Anwendungen relevanten Funktionen effizient berechnet werden können. Dies sind die Zielfunktionen der Optimierungsprobleme $(\text{LS}_{\hat{\theta}})$, (KQ1) und (KQ2) , welche in Kapitel 3 theoretisch behandelt werden und dazu dienen, „Risk-Budgeting“-Portfolios zu berechnen:

$$\begin{aligned}
 (\text{KQ1}) \quad & \min_{w \in \mathbb{R}^n} f_1(w) \\
 & s.t. \quad Aw \geq \alpha \\
 & \quad \quad Bw = \beta
 \end{aligned}$$

$$\begin{aligned}
 (\text{KQ2}) \quad & \min_{\substack{w \in \mathbb{R}^n \\ c \in \mathbb{R}}} f_2(w, c) \\
 & s.t. \quad Aw \geq \alpha \\
 & \quad \quad Bw = \beta
 \end{aligned}$$

$$\begin{aligned}
 (\text{LS}_{\hat{\theta}}) \quad & \min_{w \in \mathbb{R}^n} f_{\text{LS}}^{\hat{\theta}}(w) \\
 & s.t. \quad \theta_i w_i \geq 0, \quad i = 1, \dots, n
 \end{aligned}$$

5. Portfoliooptimierung in der Praxis

Die entsprechenden Zielfunktionen lauten wie folgt:

$$\begin{aligned}
 f_{\text{LS}}^{\hat{c}} : \mathbb{R}^n &\rightarrow \mathbb{R}, & w &\mapsto \frac{1}{2} w^\top \Sigma w - \hat{c} \sum_{i=1}^n b_i \ln \theta_i w_i \\
 f_1 : \mathbb{R}^n &\rightarrow \mathbb{R}, & w &\mapsto \sum_{i=1}^n b_i \left(\frac{w_i (\Sigma w)_i}{b_i} - w^\top \Sigma w \right)^2 \\
 f_2 : \mathbb{R}^n \times \mathbb{R} &\rightarrow \mathbb{R}, & (w, c) &\mapsto \sum_{i=1}^n b_i \left(\frac{w_i (\Sigma w)_i}{b_i} - c \right)^2
 \end{aligned}$$

Dabei sind $\theta \in \{-1, 1\}^n$, $\hat{c} > 0$, $A \in \mathbb{R}^{p \times n}$, $\alpha \in \mathbb{R}^p$, $B \in \mathbb{R}^{q \times n}$, $\beta \in \mathbb{R}^q$ und $\Sigma \in \mathbb{R}^{n \times n}$ ist eine (symmetrische, positiv semidefinite) Kovarianzmatrix. Darüber hinaus kann es auch nützlich sein, in $f_2(w, c)$, die Variable c mit $c = \hat{c} > 0$ zu fixieren (vgl. die diesbezüglichen Bemerkungen in Abschnitt 3.3.2). Die resultierende Funktion nennen wir $f_3^{\hat{c}}(w)$:

$$f_3^{\hat{c}} : \mathbb{R}^n \rightarrow \mathbb{R}, \quad w \mapsto \sum_{i=1}^n b_i \left(\frac{w_i (\Sigma w)_i}{b_i} - \hat{c} \right)^2$$

Wir versuchen nun, die benötigten Gradienten und Hesse-Matrizen von $f_{\text{LS}}^{\hat{c}}$, f_1 , f_2 und $f_3^{\hat{c}}$ händisch zu ermitteln. Um unserem iterativen SQP-Algorithmus besonders kompakte Ausdrücke zur Verfügung stellen zu können und keine expliziten Aufsummierungen implementieren zu müssen, ist es unser Ziel, Darstellungen ohne die Verwendung von Summen anzugeben. Hierfür werden neben den Notationen, die wir zu Beginn von Abschnitt 3.3.1 definiert haben, noch die nachstehenden Schreibweisen nützlich sein:

$$\begin{aligned}
 D_w &:= \text{diag}(w_1, \dots, w_n) \\
 D_b &:= \text{diag}(b_1, \dots, b_n) \\
 D_{\Sigma w} &:= \text{diag}((\Sigma w)_1, \dots, (\Sigma w)_n) \\
 M_{\Sigma w} &:= (\Sigma w, \dots, \Sigma w)
 \end{aligned}$$

$M_{\Sigma w}$ bezeichnet also diejenige $\mathbb{R}^{n \times n}$ -Matrix, für die jede Spalte gleich Σw ist.

Zuerst geben wir die Resultate für das Programm $(\text{LS}_{\hat{\theta}}^{\hat{c}})$ mit logarithmischer Schranke an:

Satz 5.1.1. *Für den Gradienten und die Hesse-Matrix der Funktion $f_{\text{LS}}^{\hat{c}}(w)$ mit symmetrischer, positiv semidefiniter Matrix $\Sigma \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}_{+,1}^n$, $\theta \in \{-1, 1\}^n$ und einem konstanten Skalar $\hat{c} > 0$ gelten für alle $w \in \mathbb{R}_{\hat{\theta}}^n$ folgende Darstellungen:*

$$\begin{aligned}
 \nabla_w f_{\text{LS}}^{\hat{c}}(w) &= \Sigma w - \hat{c} \cdot D_w^{-1} b \\
 \nabla_{ww}^2 f_{\text{LS}}^{\hat{c}}(w) &= \Sigma + \hat{c} \cdot D_w^{-2} D_b
 \end{aligned}$$

Für die Kleinste-Quadrate-Programme können wir folgenden Satz formulieren:

5. Portfoliooptimierung in der Praxis

Satz 5.1.2. Für die Gradienten bzw. Hesse-Matrizen der drei Funktionen $f_1(w)$, $f_2(w, c)$ und $f_3^{\hat{c}}(w)$ mit symmetrischer, positiv semidefiniter Matrix $\Sigma \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}_{+,1}^n$ und einem konstanten Skalar $\hat{c} > 0$ gelten für alle $w \in \mathbb{R}^n$ und $c \in \mathbb{R}$ folgende Darstellungen:

$$\begin{aligned}
 \text{a)} \quad & \nabla_w f_1(w) = 2 \cdot \Sigma D_w D_{\Sigma w} D_b^{-1} w + 2 \cdot D_{\Sigma w}^2 D_b^{-1} w - 4w^\top \Sigma w \cdot \Sigma w \\
 & \nabla_{ww}^2 f_1(w) = 2 \cdot \Sigma D_w^2 D_b^{-1} \Sigma + 4 \cdot \Sigma D_w D_{\Sigma w} D_b^{-1} \\
 & \quad + 4 \cdot D_w D_{\Sigma w} D_b^{-1} \Sigma + 2 \cdot D_{\Sigma w}^2 D_b^{-1} - 4w^\top \Sigma w \cdot \Sigma - 8 \cdot M_{\Sigma w} D_{\Sigma w} \\
 \text{b)} \quad & \nabla_w f_2(w, c) = 2 \cdot \Sigma D_w D_{\Sigma w} D_b^{-1} w + 2 \cdot D_{\Sigma w}^2 D_b^{-1} w - 4c \cdot \Sigma w \\
 & \partial / \partial c f_2(w, c) = -2 \cdot w^\top \Sigma w + 2c \\
 & \nabla_{ww}^2 f_2(w, c) = 2 \cdot \Sigma D_w^2 D_b^{-1} \Sigma + 4 \cdot \Sigma D_w D_{\Sigma w} D_b^{-1} \\
 & \quad + 4 \cdot D_w D_{\Sigma w} D_b^{-1} \Sigma + 2 \cdot D_{\Sigma w}^2 D_b^{-1} - 4c \cdot \Sigma \\
 & \nabla_w (\partial / \partial c f_2(w, c)) = -4 \cdot \Sigma w \\
 & \partial^2 / \partial c^2 f_2(w, c) = 2 \\
 \text{c)} \quad & \nabla_w f_3^{\hat{c}}(w) = 2 \cdot \Sigma D_w D_{\Sigma w} D_b^{-1} w + 2 \cdot D_{\Sigma w}^2 D_b^{-1} w - 4\hat{c} \cdot \Sigma w \\
 & \nabla_{ww}^2 f_3^{\hat{c}}(w) = 2 \cdot \Sigma D_w^2 D_b^{-1} \Sigma + 4 \cdot \Sigma D_w D_{\Sigma w} D_b^{-1} \\
 & \quad + 4 \cdot D_w D_{\Sigma w} D_b^{-1} \Sigma + 2 \cdot D_{\Sigma w}^2 D_b^{-1} - 4\hat{c} \cdot \Sigma
 \end{aligned}$$

Beweise dieser beiden Sätze sind in der Projektarbeit nachzulesen. Dort wurde auch anhand numerischer Tests gezeigt, dass die direkte Berechnung von Gradienten und Hesse-Matrizen mithilfe der Ausdrücke in Satz 5.1.1 bzw. 5.1.2 deutlich schneller möglich ist als die Verwendung entsprechender numerischer Verfahren, die vom *Octave-Forge*-Paket `optim` zur Verfügung gestellt werden.

Damit ist es uns nun möglich, bei Verwendung des globalen SQP-Algorithmus in jeder Iteration sehr effizient die Hesse-Matrix im aktuellen Iterationspunkt zu erzeugen. Wenn diese nicht positiv definit ist und daher die Matrix aus der BFGS-Formel zur Anwendung kommt, war die Berechnung zwar umsonst, der zusätzliche Zeitaufwand ist jedoch vernachlässigbar.

5.2. Numerische Ergebnisse

Wir möchten nun in einem ersten Schritt anhand realer Daten versuchen, mithilfe der implementierten Algorithmen und der nichtlinearen Optimierungsprobleme, welche im vorhergehenden Abschnitt angeführt sind, „Risk-Parity“-Portfolios zu berechnen. Ziel ist es, zu überprüfen, ob die genannten Programme in der Praxis handhabbar und mit unserem globalen SQP-Algorithmus lösbar sind. Vor allem die Tatsache, dass die Zielfunktionen f_1 , f_2 und $f_3^{\hat{c}}$ nicht konvex sind, sollte dabei nicht außer Acht gelassen werden.

In einem zweiten Schritt werden wir unser „Risk-Parity“-Modell mit dem grundlegenden Markowitz’schen Modell, welches im 2. Kapitel vorgestellt wird, vergleichen.

5. Portfoliooptimierung in der Praxis

Wie bereits implizit angedeutet, beschränken wir uns bei unseren Untersuchungen auf den „Risk-Parity“-Spezialfall unseres „Risk-Budgeting“-Modells. Wir setzen also bei der Definition des Vektors $b \in \mathbb{R}^n$, für den im allgemeinen „Risk-Budgeting“-Problem $b_i > 0, \forall i = 1, \dots, n$ und $\sum_{i=1}^n b_i = 1$ erfüllt sein muss, stets $b_i = 1/n$. Gleich hohe Risikoanteile für alle Aktien erscheinen uns nämlich zur Veranschaulichung der risikobasierten Portfoliooptimierung am geeignetsten.

Wir werden einen amerikanischen Aktienindex, nämlich den *Down Jones Composite Average* (DJCA), verwenden. Dieser beinhaltet die Aktien von 65 der bedeutendsten US-amerikanischen Unternehmen und fasst die folgenden drei Indizes zusammen:

- *Down Jones Industrial Average* (DJIA),
Aktien von 30 der bedeutendsten US-amerikanischen Unternehmen
- *Down Jones Transportation Average* (DJTA),
Aktien von 20 US-amerikanischen Transportunternehmen
- *Down Jones Utility Average* (DJUA),
Aktien von 15 US-amerikanischen Versorgungsunternehmen

Dabei ist zu beachten, dass der DJIA nicht nur Aktien von Industrieunternehmen beinhaltet. Die Bezeichnung „Industrial“ ist historisch bedingt.

Sowohl der DJCA als auch seine drei Subindizes sind preisgewichtet. Der Wert eines solchen Index ergibt sich aus der Summe der Kurse aller enthaltenen Aktien, dividiert durch einen Divisor. Dieser Divisor, welchen man für den DJIA auch „Dow Divisor“ nennt, wird beispielsweise bei Aktiensplits, nicht jedoch bei Dividendenzahlungen, angepasst. Dies geschieht um sicherzustellen, dass der Kurs des Aktienindex konsistent bleibt.

Die entsprechenden Aktienkurse entnehmen wir aus *Yahoo Finance* (siehe <http://finance.yahoo.com/>). Dort sind nämlich (im Gegensatz zu *Google Finance*) sogenannte „Adjusted Close“-Preise vorhanden, die eine um Dividendenzahlungen und Aktiensplits bereinigte Version der täglichen Aktien-Schlusskurse angeben (vgl. <https://help.yahoo.com/kb/finance/historical-prices-sln2311.html>). Wir verwenden die Zusammensetzung des DJCA aus dem Jahr 2013, welche in der nachstehenden Tabelle ersichtlich ist.

#	Index	Yahoo-Symbol	Unternehmen	Daten verfügbar ab
1	DJIA	AA	Alcoa Inc.	02.01.1962
2	DJIA	AXP	American Express Company	01.06.1972
3	DJIA	BA	The Boeing Company	02.01.1962
4	DJIA	BAC	Bank of America Corporation	29.05.1986
5	DJIA	CAT	Caterpillar Inc.	02.01.1962
6	DJIA	CSCO	Cisco Systems Inc.	26.03.1990
7	DJIA	CVX	Chevron Corporation	02.01.1970
8	DJIA	DD	E. I. du Pont de Nemours and Company	02.01.1962
9	DJIA	DIS	The Walt Disney Company	02.01.1962
10	DJIA	GE	General Electric Company	02.01.1962
11	DJIA	HD	The Home Depot Inc.	22.09.1981

5. Portfoliooptimierung in der Praxis

12	DJIA	HPQ	Hewlett-Packard Company	02.01.1962
13	DJIA	IBM	International Business Machines Corporation	02.01.1962
14	DJIA	INTC	Intel Corporation	17.03.1980
15	DJIA	JNJ	Johnson & Johnson	02.01.1970
16	DJIA	JPM	JPMorgan Chase & Co.	30.12.1983
17	DJIA	KO	The Coca-Cola Company	02.01.1962
18	DJIA	MCD	McDonald's Corp.	02.01.1970
19	DJIA	MMM	3M Company	02.01.1970
20	DJIA	MRK	Merck & Co. Inc.	02.01.1970
21	DJIA	MSFT	Microsoft Corporation	13.03.1986
22	DJIA	PFE	Pfizer Inc.	01.06.1972
23	DJIA	PG	The Procter & Gamble Company	02.01.1970
24	DJIA	T	AT&T Inc.	19.07.1984
25	DJIA	TRV	The Travelers Companies Inc.	09.07.1986
26	DJIA	UNH	UnitedHealth Group Incorporated	26.03.1990
27	DJIA	UTX	United Technologies Corporation	02.01.1970
28	DJIA	VZ	Verizon Communications Inc.	21.11.1983
29	DJIA	WMT	Wal-Mart Stores Inc.	25.08.1972
30	DJIA	XOM	Exxon Mobil Corporation	02.01.1970
31	DJTA	ALK	Alaska Air Group Inc.	03.01.1983
32	DJTA	CHRW	CH Robinson Worldwide Inc.	16.10.1997
33	DJTA	CNW	Con-way Inc.	02.01.1980
34	DJTA	CSX	CSX Corp.	03.11.1980
-	DJTA	DAL	Delta Air Lines Inc.	03.05.2007
35	DJTA	EXPD	Expeditors International of Washington Inc.	26.03.1990
36	DJTA	FDX	FedEx Corporation	12.04.1978
37	DJTA	GMT	GATX Corp.	05.11.1987
38	DJTA	JBHT	JB Hunt Transport Services Inc.	26.03.1990
39	DJTA	JBLU	JetBlue Airways Corporation	12.04.2002
40	DJTA	KEX	Kirby Corporation	21.08.1991
41	DJTA	KSU	Kansas City Southern	05.11.1987
42	DJTA	LSTR	Landstar System Inc.	05.03.1993
43	DJTA	LUV	Southwest Airlines Co.	02.01.1980
44	DJTA	MATX	Matson Inc.	03.05.1973
45	DJTA	NSC	Norfolk Southern Corporation	02.06.1982
46	DJTA	R	Ryder System Inc.	02.01.1980
-	DJTA	UAL	United Continental Holdings Inc.	06.02.2006
47	DJTA	UNP	Union Pacific Corporation	02.01.1980
48	DJTA	UPS	United Parcel Service Inc.	10.11.1999
49	DJUA	AEP	American Electric Power Co. Inc.	02.01.1970
50	DJUA	AES	The AES Corporation	26.06.1991
51	DJUA	CNP	CenterPoint Energy Inc.	02.01.1970
52	DJUA	D	Dominion Resources Inc.	03.10.1984
53	DJUA	DUK	Duke Energy Corporation	06.04.1983
54	DJUA	ED	Consolidated Edison Inc.	02.01.1970
55	DJUA	EIX	Edison International	02.01.1980
56	DJUA	EXC	Exelon Corporation	02.01.1980
57	DJUA	FE	FirstEnergy Corp.	10.11.1997
58	DJUA	NEE	NextEra Energy Inc.	10.06.1983
59	DJUA	NI	NiSource Inc.	19.10.1984
60	DJUA	PCG	PG&E Corporation	01.06.1972
61	DJUA	PEG	Public Service Enterprise Group Inc.	02.01.1980
62	DJUA	SO	Southern Company	31.12.1981
63	DJUA	WMB	Williams Companies Inc.	31.12.1981

Tab. 5.1.: Aktien des DJCA sowie seiner Teilindizes DJIA, DJTA und DJUA

5. Portfoliooptimierung in der Praxis

Mit Ausnahme der beiden gelb markierten Unternehmen sind sämtliche Aktienkurse mindestens seit 12. April 2002 verfügbar. Wir werden für die weiteren Betrachtungen diese zwei Aktien ignorieren und mit den restlichen 63 (statt 65) Aktien des DJCA bzw. 18 (statt 20) Aktien des DJTA weiterarbeiten. Damit wird es uns möglich sein, verschiedene Zeiträume vor und während der im Sommer 2007 beginnenden Finanzkrise zu vergleichen.²⁷

Wir möchten nun die täglichen „Adjusted Close“-Preise dieser 63 Aktien aus *Yahoo Finance* für verschiedene Zeiträume einlesen und daraus die täglichen Returns ermitteln. Dafür wurde auf die im *Octave-Forge*-Paket `financial` enthaltene Funktion `fetch` zurückgegriffen, mit der die Aktienkurse direkt aus *Yahoo Finance* gezogen werden können.

Bei der Ermittlung der Returns halten wir uns an die Definition, welche wir im Zusammenhang mit dem grundlegenden Markowitz'schen Modell in Kapitel 2 verwendet haben: Der Return der i -ten Aktie in einem Zeitraum $[T_0, T_1]$ ist demnach eine Zufallsvariable und wird durch

$$R_i = \frac{P_i^{T_1} - P_i^{T_0}}{P_i^{T_0}}$$

beschrieben, wobei $P_i^{T_0}$ dem (bekannten) Preis des i -ten Assets zum Zeitpunkt T_0 und $P_i^{T_1}$ dem (unbekannten, zufälligen) Preis des i -ten Assets zum Zeitpunkt T_1 entspricht.

Für die Berechnung der realisierten Returns ziehen wir (wie bereits erwähnt) die „Adjusted Close“-Preise heran. Diese multiplizieren wir jedoch anschließend noch mit 100, um die prozentualen Werte, also die Realisierungen der Zufallsvariablen $\tilde{R}_i := R_i \cdot 100$, zu erhalten. Dies hat in erster Linie numerische Gründe, da die täglichen Returns und damit auch deren Erwartungswerte und Varianzen naturgemäß sehr klein sind.

Für die Untersuchungen der folgenden beiden Unterkapitel benötigen wir nun Schätzungen $\hat{\Sigma}$ der Kovarianzmatrix Σ der zufälligen Returns ($\sigma_{i,j} = \text{Cov}(R_i, R_j)$) und für die vergleichende Analyse von „Risk-Parity“ und „Mean-Variance“ auch Schätzungen $\hat{\mu}$ der entsprechenden Erwartungswerte μ (mit $\mu_i = \mathbb{E}(R_i)$). Diese Werte werden stets durch die Verwendung der in *GNU Octave* vorhandenen Funktionen `cov` bzw. `mean` berechnet. Weitreichendere Schätzmethoden werden im Rahmen dieser Arbeit nicht behandelt.

5.2.1. Berechnung von „Risk-Parity“-Portfolios

Wir beginnen nun damit, zu überprüfen, ob es möglich ist, mithilfe der vier Funktionen f_1 , f_2 , $f_3^{\hat{c}}$ sowie $f_{LS}^{\hat{c}}$ die gesuchten „Risk-Parity“-Portfolios zu ermitteln. In diesem Abschnitt werden wir den *Dow Jones Composite Average* mit den angegebenen $n = 63$ Aktien verwenden. Die Schätzung $\hat{\Sigma}$ der Kovarianzmatrix Σ ermitteln wir aus den täglichen Returns von 1. Januar 2013 bis 31. Dezember 2014. Diese Zeitspanne entspricht ca. 500

²⁷Es ist zu berücksichtigen, dass der DJCA in seiner Zusammensetzung im Jahr 2014 noch zwei weitere Aktien beinhaltet, deren Kurse in *Yahoo Finance* erst seit dem Jahr 2008 zur Verfügung gestellt werden. Daher verwenden wir auch die Indexkomponenten von 2013.

5. Portfoliooptimierung in der Praxis

Datenpunkten. Die Wahl des Zeitraumes ist hier jedoch nebensächlich, da es uns vorerst nur um die prinzipielle Lösbarkeit des „Risk-Parity“-Problems mithilfe der erwähnten Optimierungsprobleme geht.

Wir wissen aus den theoretischen Überlegungen in Abschnitt 3.3.1, dass es höchstens ein „Risk-Parity“-Portfolio gibt, falls wir für jedes Asset das Vorzeichen des entsprechenden Assetgewichts vorgeben und die Summe aller Assetgewichte auf 1 normieren (siehe Korollar 3.3.6). Folgende vier Optimierungsprobleme werden im Folgenden gelöst, wobei die Zielfunktionen wieder jenen aus Kapitel 5.1 (mit $b_i = 1/n, \forall i = 1, \dots, n$ und $\Sigma = \hat{\Sigma}$) entsprechen:

$$\begin{array}{ll}
 \widehat{\text{KQ1}} & \min_{w \in \mathbb{R}^{63}} f_1(w) \\
 & s.t. \quad \theta_i w_i \geq 0, \quad i = 1, \dots, 63 \\
 & \quad \sum_{i=1}^{63} w_i = 1 \\
 \widehat{\text{KQ2}} & \min_{\substack{w \in \mathbb{R}^{63} \\ c \in \mathbb{R}}} f_2(w, c) \\
 & s.t. \quad \theta_i w_i \geq 0, \quad i = 1, \dots, 63 \\
 & \quad \sum_{i=1}^{63} w_i = 1
 \end{array}$$

$$\begin{array}{ll}
 \widehat{\text{KQ3}}^{\hat{c}} & \min_{w \in \mathbb{R}^{63}} f_3^{\hat{c}}(w) \\
 & s.t. \quad \theta_i w_i \geq 0, \quad i = 1, \dots, 63 \\
 \text{LS}_{\hat{\theta}}^{\hat{c}} & \min_{w \in \mathbb{R}^{63}} f_{\text{LS}}^{\hat{c}}(w) \\
 & s.t. \quad \theta_i w_i \geq 0, \quad i = 1, \dots, 63
 \end{array}$$

Es ist zu beachten, dass in den Optimierungsproblemen $\widehat{\text{KQ1}}$ und $\widehat{\text{KQ2}}$ nicht nur die Vorzeichen der Assetgewichte als Ungleichungen fixiert sind, sondern auch unmittelbar die Summe der Gewichte in Form einer Gleichheitsrestriktion gleich 1 gesetzt werden kann. Für $\text{LS}_{\hat{\theta}}^{\hat{c}}$ und $\widehat{\text{KQ3}}^{\hat{c}}$ wissen wir aus Abschnitt 3.3.1 bzw. 3.3.2, dass dies nicht möglich ist. In diesen beiden Programmen wird stattdessen der konstante Wert $\hat{c} > 0$ beliebig gewählt und nach Lösung der Optimierungsprobleme eine Gewichtsnormierung vorgenommen. Wir verwenden hier sowohl für $\text{LS}_{\hat{\theta}}^{\hat{c}}$ als auch für $\widehat{\text{KQ3}}^{\hat{c}}$ stets $\hat{c} = 1$.

$\widehat{\text{KQ1}}$ und $\widehat{\text{KQ2}}$ sind auch die einzigen beiden Formulierungen, die noch sinnvoll um weitere Gleichheits- und Ungleichheitsrestriktionen (wie obere und untere Schranken für die Gewichte oder Bedingungen an den Return) erweitert werden können. Unser Hauptaugenmerk liegt also auf diesen beiden praxisrelevanten Programmen.

Wir führen nun sämtliche Berechnungen für zwei verschiedene Vektoren $\theta \in \{-1, 1\}^{63}$ (also für verschiedene Gewichtsvorzeichen) durch:

- **1. Test:** $\theta_i = 1, i \in \{1, \dots, 63\}$
Zuerst interessieren wir uns also für den „Long-Only“-Fall.
- **2. Test:** $\theta_i = -1, i \in \{1, 31, 49\}$ und $\theta_i = 1, i \in \{1, \dots, 63\} \setminus \{1, 31, 49\}$
Zum Vergleich versuchen wir, ein „Risk-Parity“-Portfolio zu ermitteln, in dem auch „Short Selling“ erlaubt ist. Hierfür setzen wir jeweils für eine Aktie der drei Teilindizes des DJCA, nämlich für „Alcoa Inc.“ aus dem *Dow Jones Industrial Average* ($i = 1$),

5. Portfoliooptimierung in der Praxis

„Alaska Air Group Inc.“ aus dem *Dow Jones Transportation Average* ($i = 31$) und „American Electric Power Co. Inc.“ aus dem *Dow Jones Utility Average* ($i = 49$), das Vorzeichen auf -1 . Die Wahl dieser drei Aktien ist willkürlich und entspricht für jeden Subindex jeweils der alphabetisch ersten Komponente. In den übrigen 60 Aktien streben wir wieder „Long Positions“ an.

Wir wenden unsere Implementierung des globalen SQP-Algorithmus auf unsere vier Optimierungsprobleme an, wobei wir zur Ermittlung des jeweiligen Gradienten und der Hesse-Matrix in jeder Iteration auf die Resultate der Sätze 5.1.1 und 5.1.2 zurückgreifen. Als zulässigen Startpunkt wählen wir den Vektor w^0 mit $w_i^0 = \frac{\theta_i}{\sum_{i=1}^{63} \theta_i}$ und in f_2 zusätzlich $c^0 = (w^0)^\top \Sigma w^0$ als Startwert für die Variable c . Dies entspricht im „Long-Only“-Fall $w_i = 1/63$ („Weight-Budgeting“, siehe Abschnitt 3.3.1.1). Im zweiten Test gilt $w_i = \theta_i/57$. Für die Lösung der konvexen, quadratischen Teilprobleme, die in jeder Iteration des SQP-Algorithmus auftreten, verwenden wir pro Optimierungsproblem einmal unsere Implementierung der Methode der aktiven Menge und einmal unser unzulässiges Innere-Punkte-Verfahren nach Mehrotra, um auch einen Vergleich zwischen diesen beiden Varianten ziehen zu können. Wir interessieren uns für die Laufzeit der Algorithmen, die Anzahl der Iterationen (wobei wir auch berücksichtigen, wie oft die Matrix aus der BFGS-Formel bzw. die Hesse-Matrix der Zielfunktion im aktuellen Iterationspunkt verwendet wird), die größten und kleinsten totalen Risikobeiträge in den ermittelten Portfolios (woraus direkt ersichtlich ist, ob es sich um eine „Risk-Parity“-Lösung handelt), und die Portfolio-Varianzen. Die Ergebnisse sind in den nachstehenden beiden Tabellen aufgelistet:

Test 1: „Long-Only“, $\theta_i = 1, i \in \{1, \dots, 63\}$							
Probl.	Methode	Laufzeit	BFGS	Hesse	min. ζ_i^2	max. ζ_i^2	PF-Var.
$(\widehat{KQ1})$	Akt. Men.	19.757 Sek.	97	0	0.00647	0.00647	0.407
$(\widehat{KQ1})$	Inn. Pkt.	101.010 Sek.	97	0	0.00647	0.00647	0.407
$(\widehat{KQ2})$	Akt. Men.	1.473 Sek.	4	4	0.00647	0.00647	0.407
$(\widehat{KQ2})$	Inn. Pkt.	7.874 Sek.	4	4	0.00647	0.00647	0.407
$(\widehat{KQ3^c})$	Akt. Men.	0.092 Sek.	0	6	0.00647	0.00647	0.407
$(\widehat{KQ3^c})$	Inn. Pkt.	5.057 Sek.	0	6	0.00647	0.00647	0.407
(\widehat{LS}_θ^c)	Akt. Men.	0.059 Sek.	0	5	0.00647	0.00647	0.407
(\widehat{LS}_θ^c)	Inn. Pkt.	4.216 Sek.	0	5	0.00647	0.00647	0.407

Tab. 5.2.: Berechnung eines „Risk-Parity“-Portfolios im „Long-Only“-Fall für den DJCA

5. Portfoliooptimierung in der Praxis

Test 2: $\theta_i = -1, i \in \{1, 31, 49\}$ und $\theta_i = 1, i \in \{1, \dots, 63\} \setminus \{1, 31, 49\}$							
Probl.	Methode	Laufzeit	BFGS	Hesse	min. ζ_i^2	max. ζ_i^2	PF-Var.
$\widehat{(\text{KQ1})}$	Akt. Men.	24.407 Sek.	99	0	≈ 0	0.00698	0.398
$\widehat{(\text{KQ1})}$	Inn. Pkt.	117.527 Sek.	99	0	≈ 0	0.00698	0.398
$\widehat{(\text{KQ2})}$	Akt. Men.	4.491 Sek.	90	1	≈ 0	0.00698	0.398
$\widehat{(\text{KQ2})}$	Inn. Pkt.	102.237 Sek.	90	1	≈ 0	0.00698	0.398
$\widehat{(\text{KQ3}^{\hat{c}})}$	Akt. Men.	0.129 Sek.	1	4	≈ 0	0.00668	0.401
$\widehat{(\text{KQ3}^{\hat{c}})}$	Inn. Pkt.	4.197 Sek.	1	4	≈ 0	0.00668	0.401
$(\text{LS}_{\hat{\theta}}^{\hat{c}})$	Akt. Men.	0.081 Sek.	0	8	0.01226	0.01226	0.772
$(\text{LS}_{\hat{\theta}}^{\hat{c}})$	Inn. Pkt.	6.673 Sek.	0	8	0.01225	0.01226	0.772

Tab. 5.3.: Berechnung eines „Risk-Parity“-Portfolios mit „Long“ und „Short Positions“ für den DJCA

Wir vergleichen nun die einzelnen Ergebnisse und machen folgende Beobachtungen:

- Sowohl die Implementierung der Methode der aktiven Menge als auch des „Predictor-Corrector“-Verfahrens nach Mehrotra sind in der Lage, die iterativ auftretenden quadratischen Teilprobleme zu lösen. Allerdings sind die Laufzeiten bei Verwendung des ersteren Lösungsverfahrens um ein vielfaches geringer. Es ist jedoch stark anzunehmen, dass dies durch den Einsatz von effizienteren (gegebenenfalls numerischen) Methoden für die Lösung der Gleichungssysteme (4.21) und (4.23), die im „Predictor-Corrector“-Verfahren auftreten, geändert werden kann. Darüber hinaus ist zu berücksichtigen, dass wir für die Innere-Punkte-Methode als Input quadratische Programme der Form $\widehat{(\text{QP})}$ benötigen (siehe Kapitel 4.1.3). Die entsprechende Transformation von (QP) auf $\widehat{(\text{QP})}$, die in Kapitel 1.3 erläutert wird, führt jedoch zu einer deutlichen Erhöhung der Dimension (von n auf $\tilde{n} = 2n + p$ mit p für die Anzahl der Ungleichheitsrestriktionen).
- Die berechneten lokalen Optima des konvexen Optimierungsproblems $(\text{LS}_{\hat{\theta}}^{\hat{c}})$ mit logarithmischer Schranke sind klarerweise auch globale Optima und damit die gesuchten „Risk-Parity“-Lösungen. Dies erkennen wir auch am Vergleich von größtem und kleinstem Risikobeitrag, welche in beiden Tests übereinstimmen. Die Varianz des „Long-Only“-Portfolios im „Risk-Parity“-Zustand ist mit 0.407 deutlich kleiner als für den zweiten Fall, in dem auch drei „Short Positions“ vorkommen (0.772). Wie zu erwarten, wurde aufgrund der Konvexität von $f_{\text{LS}}^{\hat{c}}$ stets die positiv definite Hesse-Matrix anstatt der BFGS-Matrix genommen.
- Für $\widehat{(\text{KQ1})}$ war es hingegen nie möglich, auf die Hesse-Matrix zurückzugreifen. Bei $\widehat{(\text{KQ2})}$ treten sowohl Iterationen mit der Hesse- als auch mit der BFGS-Matrix auf. In $\widehat{(\text{KQ3}^{\hat{c}})}$ überwiegen sogar die Iterationen mit Hesse-Matrizen deutlich.
- Im „Long-Only“-Fall sind alle berechneten Lösungen globale Lösungen. Sämtliche Zielfunktionswerte für f_1, f_2 und $f_3^{\hat{c}}$ sind kleiner als 10^{-8} , die kleinsten und größten Risikobeiträge stimmen überein und die Lösungen sind damit „Risk-Parity“-Portfolios.

5. Portfoliooptimierung in der Praxis

Es hat also keine Probleme bereitet, dass diese drei Funktionen nicht konvex sind. Auffallend ist, dass die Lösung von f_2 deutlich schneller erfolgt als von f_1 . Es erscheint also tatsächlich auch für praktische Anwendungen (und nicht nur für unsere theoretischen Betrachtungen in Kapitel 3.3.2) sinnvoll, den quadratischen Term $w^\top \Sigma w$ in f_1 durch die zusätzliche Variable c zu ersetzen.

- Der zweite Test zeigt ein etwas anderes Bild: Mit dem gewählten Startpunkt sind wir bei allen drei nicht-konvexen Funktionen in lokalen Optima gelandet, die nicht global optimal sind. Dies erkennen wir daran, dass die kleinsten Risikobeiträge gleich 0 sind und nicht mit den größten Risikobeiträgen übereinstimmen. Eine genauere Betrachtung der berechneten Punkte zeigt sogar, dass die Gewichte aller drei Aktien, für die wir *negative* Gewichtsvorzeichen definiert haben, stets gleich 0 sind (und damit am Rand der Zulässigkeitsmenge liegen). Des Weiteren sehen wir, dass die Anzahl der Iterationen zur Lösung von $(\widehat{\text{KQ2}})$ im Gegensatz zum ersten Test beinahe auf dem Niveau von $(\widehat{\text{KQ1}})$ liegt. Trotzdem bietet $(\widehat{\text{KQ2}})$ vor allem bei Verwendung der Methode der aktiven Menge noch einen deutlichen Laufzeitvorteil gegenüber $(\widehat{\text{KQ1}})$.
- Die Algorithmen terminieren für die Zielfunktion f_3^c mit festem \hat{c} teils deutlich schneller als für f_2 mit der Variable c . Doch wie bereits erwähnt, können wir dies nur für bestimmte Spezialfälle (keine Restriktionen außer Vorzeichenbedingungen) nutzen.

Um als Abschluss dieser ersten praktischen Analysen noch einen empirischen Vergleich ziehen zu können, in welchem Verhältnis die benötigten Iterationen unserer beiden Algorithmen zur Lösung konvexer quadratischer Probleme stehen, betrachten wir noch kurz folgendes „Minimum-Variance“-Programm, wobei wir wieder den *Dow Jones Composite Average* zugrunde legen:

$$\begin{aligned}
 (\widehat{\text{MV}}) \quad & \min_{w \in \mathbb{R}^{63}} \quad \frac{1}{2} w^\top \hat{\Sigma} w \\
 & \text{s.t.} \quad w_i \geq 0, \quad i = 1, \dots, 63 \\
 & \quad \sum_{i=1}^{30} w_i \geq 1/5, \quad \sum_{i=31}^{48} w_i \geq 1/5, \quad \sum_{i=49}^{63} w_i \geq 1/5 \\
 & \quad \sum_{i=1}^{63} w_i = 1
 \end{aligned}$$

Wir nehmen also eine Gewichtsnormierung vor und verlangen, dass die Gewichtssumme der Aktien der drei Subindizes DJIA ($i = 1, \dots, 30$), DJTA ($i = 31, \dots, 48$) und DJUA ($i = 49, \dots, 63$) jeweils mindestens 20% beträgt. Wir erhalten folgende Resultate:

Test 3: „Minimum-Variance“					
Methode	Laufzeit	Iterationen	min. ζ_i^2	max. ζ_i^2	PF-Varianz
Aktive Menge	0.790 Sek.	46	≈ 0	0.0480	0.283
Innere Punkte	1.285 Sek.	9	≈ 0	0.0480	0.283

Tab. 5.4.: Berechnung eines „Minimum-Variance“-Portfolios für den DJCA

5. Portfoliooptimierung in der Praxis

Wir sehen, dass das unzulässige Innere-Punkte-Verfahren deutlich weniger Iterationen benötigt als die Methode der aktiven Menge. Effizientere Lösungsmethoden für die (vergleichsweise) aufwendigen Gleichungssysteme (4.21) und (4.23) in Mehrotras „Predictor-Corrector“-Verfahren lassen also tatsächlich Laufzeitverbesserungen erhoffen, was nicht mehr im Rahmen dieser Arbeit liegt. Wir haben in unserer Umsetzung dieser Methode nur berücksichtigt, dass (4.21) und (4.23) über die gleiche Koeffizientenmatrix verfügen, und daher eine LR-Zerlegung verwendet, um die Matrixfaktorisierung nur ein Mal pro Iteration durchführen zu müssen und damit Laufzeit zu sparen. Die Funktionalität der implementierten Algorithmen wurde hiermit jedoch gezeigt.

5.2.2. Eine vergleichende Analyse von „Risk-Parity“ und „Mean-Variance“

In diesem Kapitel möchten wir nun sowohl für den *Dow Jones Composite Average* als auch für jeden der drei kleineren Teilindizes *Dow Jones Industrial Average*, *Dow Jones Transportation Average* sowie *Dow Jones Utility Average* das „Risk-Parity“-Modell und das Markowitz’sche „Mean-Variance“-Modell vergleichen. Dies ist möglich, da uns mit den besprochenen Kleinste-Quadrate-Optimierungsproblemen Formulierungen des „Risk-Parity“-Modells zur Verfügung stehen, in denen wir die selbe Zulässigkeitsmenge wie für die „Mean-Variance“-Probleme zugrunde legen können. Konkret verwenden wir nachstehende zwei Programme:

$$\begin{array}{ll}
 \widetilde{\text{(MW1)}} & \min_{w \in \mathbb{R}^n} \frac{1}{2} w^\top \Sigma w \\
 & \text{s.t. } \mu^\top w \geq r \\
 & \quad w_i \geq 0, \quad i = 1, \dots, n \\
 & \quad \sum_{i=1}^n w_i = 1 \\
 \widetilde{\text{(KQ2)}} & \min_{\substack{w \in \mathbb{R}^n \\ c \in \mathbb{R}}} f_2(w, c) \\
 & \text{s.t. } \mu^\top w \geq r \\
 & \quad w_i \geq 0, \quad i = 1, \dots, n \\
 & \quad \sum_{i=1}^n w_i = 1
 \end{array}$$

Dabei ist wieder $f_2(w, c) = \sum_{i=1}^n b_i \left(\frac{w_i (\Sigma w)_i}{b_i} - c \right)^2$ mit $b_i = 1/n$. Man beachte, dass hier der „Risk-Parity“-Ansatz nicht rein risikobasiert ist, da auch der erwartete Return eine Rolle spielt.

$\widetilde{\text{(MW1)}}$ basiert auf der Formulierung (MW1) des Markowitz’schen Modells aus dem 2. Kapitel. Ebendort sind noch zwei äquivalente Problemformulierungen angegeben. Die Entscheidung für die Verwendung der Zielfunktion $f_2(w, c)$ im „Risk-Parity“-Modell anstatt von $f_1(w) = \sum_{i=1}^n b_i \left(\frac{w_i (\Sigma w)_i}{b_i} - w^\top \Sigma w \right)^2$ treffen wir aufgrund der numerischen Resultate des vorhergehenden Abschnittes 5.2.1.

5. Portfoliooptimierung in der Praxis

Wenn wir in weiterer Folge von „*Mean-Variance*“-Portfolios sprechen, so meinen wir Lösungen von $(\widetilde{MW1})$. Die berechneten Minima von $(\widetilde{KQ2})$, von denen wir im Allgemeinen nicht wissen, ob sie nur lokale oder auch globale Optima sind, werden wir (in Übereinstimmung mit Kapitel 3.3.2) als *approximative* „*Risk-Parity*“-Portfolios bezeichnen, da ja für ausreichend große r selbst das globale Optimum von $(\widetilde{KQ2})$ nicht im „*Risk-Parity*“-Zustand ist.

Ziel wird es sein, Schätzungen von μ und Σ aus den täglichen Returns eines gewissen Zeitraumes zu berechnen, diese als Input für unsere beiden Modelle zu verwenden und anschließend die Performance der ermittelten Portfolios in einem anderen Zeitraum zu vergleichen. Wie bereits erwähnt, stehen die Daten aller 63 Aktien in *Yahoo Finance* ab 12. April 2002 vollständig zur Verfügung. Daher berechnen wir zu Beginn für jedes einzelne Jahr von 2003 bis 2014 (also mit jeweils ca. 250 Datenpunkten) eine Schätzung $\hat{\Sigma}$ der Kovarianzmatrix, auf deren Diagonale die geschätzten Varianzen $\hat{\sigma}_i^2$ der einzelnen Assets stehen, um einen Überblick über die Daten zu erhalten und geeignete Zeitperioden für unsere Vergleiche auswählen zu können. In der folgenden Tabelle sind die ermittelten Werte

$$\hat{\sigma}_{\min}^2 := \min_{i=1,\dots,n} \{\hat{\sigma}_i^2\} \quad \text{und} \quad \hat{\sigma}_{\max}^2 := \max_{i=1,\dots,n} \{\hat{\sigma}_i^2\}$$

für die Indizes DJCA, DJIA, DJTA und DJUA aufgelistet.

Jahr	DJCA		DJIA		DJTA		DJUA	
	$\hat{\sigma}_{\min}^2$	$\hat{\sigma}_{\max}^2$	$\hat{\sigma}_{\min}^2$	$\hat{\sigma}_{\max}^2$	$\hat{\sigma}_{\min}^2$	$\hat{\sigma}_{\max}^2$	$\hat{\sigma}_{\min}^2$	$\hat{\sigma}_{\max}^2$
2003	0.715	18.384	0.715	6.181	0.715	7.868	0.998	18.384
2004	0.537	7.750	0.713	4.879	0.585	7.750	0.537	4.513
2005	0.569	5.736	0.569	2.966	1.494	5.736	0.675	4.257
2006	0.455	9.155	0.455	3.202	1.483	9.155	0.535	2.824
2007	0.518	7.008	0.518	4.527	0.958	7.008	0.996	3.430
2008	3.517	39.852	3.517	39.852	6.157	36.832	3.535	23.118
2009	1.239	60.766	1.239	60.766	4.324	21.836	1.309	14.749
2010	0.655	6.845	0.655	5.516	1.785	6.845	0.679	5.295
2011	0.684	13.155	0.896	13.155	1.881	10.520	0.684	5.945
2012	0.370	6.404	0.370	6.067	0.808	6.404	0.451	2.095
2013	0.552	5.741	0.552	5.741	0.737	4.159	0.625	1.974
2014	0.544	5.715	0.544	3.628	0.725	5.715	0.700	3.581

Tab. 5.5.: Varianzen der Aktien des DJCA und seiner Teilindizes für die einzelnen Jahre von 2003 bis 2014

Wir beobachten, dass vor allem in den zwei Jahren unmittelbar nach dem Beginn der Finanzkrise, 2008 und 2009, sehr hohe Varianzen auftreten. Dagegen sind in den Jahren 2004 und 2005 die Werte für $\hat{\sigma}_{\min}^2$ und $\hat{\sigma}_{\max}^2$ empfindlich kleiner. In den Jahren 2013 und 2014 erreichen diese Kennzahlen wieder das Niveau von 2004/2005. Basierend auf diesen

5. Portfoliooptimierung in der Praxis

Erkenntnissen arbeiten wir mit den drei verschiedenen Zeiträumen von 1. Januar 2004 bis 31. Dezember 2005 (also vor der Finanzkrise), von 1. Januar 2008 bis 31. Dezember 2009 (also während der Mitte 2007 einsetzenden Finanzkrise) sowie von 1. Januar 2013 bis 31. Dezember 2014. Wir fassen also jeweils zwei vollständige, aufeinanderfolgende Jahre zusammen und verwenden somit in etwa 500 Datenpunkte pro Periode.

Wir lesen nun die täglichen Returns (in Prozent) für diese drei Perioden ein. Anschließend schätzen wir μ und Σ basierend auf dem Zeitraum 2004/2005, verwenden diese Schätzungen als Input für $(\widetilde{\text{MW1}})$ sowie $(\widetilde{\text{KQ2}})$ und lösen diese beiden Programme mit verschiedenen unteren Return-Schranken r , wobei wir 300 äquidistante Werte im Intervall $[r_{\min}, r_{\max}]$ verwenden.²⁸ Für die resultierenden Portfolios des zusammengesetzten DJCA vergleichen wir noch die realisierten Returns und Varianzen im Zeitraum der Schätzung (2004/2005) und darüber hinaus auch im zweiten Zeitraum von 1. Januar 2008 bis 31. Dezember 2009. Damit können wir feststellen, ob die approximativen „Risk-Parity“-Portfolios, die wir anhand der Return-Daten aus der Zeit vor der Finanzkrise berechnet haben, während der Finanzkrise bessere Resultate liefern als die traditionellen „Mean-Variance“-Portfolios.

Weiters werden wir noch eine zweite Schätzung aus den Daten des Zeitraumes 2008/2009 erstellen und nach dem gleichen, soeben beschriebenen Schema vorgehen. Als Vergleichszeitraum wird uns hier die Periode von 1. Januar 2013 bis 31. Dezember 2014 dienen. Wir vermuten nämlich, dass sich das „Mean-Variance“-Modell bei Schätzung von Σ aus den sehr volatilen Returns der Jahre 2008/2009 stark auf die wenigen Aktien konzentriert, welche vergleichsweise kleine Varianzen aufweisen, und aufgrund der resultierenden geringen Diversifikation wiederum „nach der Krise“ schlechter abschneidet.

Bei der Lösung von $(\widetilde{\text{MW1}})$ und $(\widetilde{\text{KQ2}})$ gehen wir wie folgt vor:

„Mean-Variance“-Problem $(\widetilde{\text{MW1}})$:

- Wir lösen die 300 konvexen quadratischen Probleme mit den verschiedenen Return-Schranken stets mit der Methode der aktiven Menge.
- Zuerst erfolgt die Lösung von $(\widetilde{\text{MW1}})$ mit $r = r_{\max}$. Anschließend verringern wir iterativ die Return-Schranke bis $r = r_{\min}$ und verwenden als Startpunkt jeweils das zuvor berechnete Portfolio, welches klarerweise (wegen $\mu^\top w \geq r$) zulässig für das nachfolgende Problem mit kleinerem r ist und hoffentlich einen guten Startwert darstellt.

„Risk-Parity“-Problem $(\widetilde{\text{KQ2}})$:

- Es kommt die Implementierung unseres globalen SQP-Algorithmus zur Anwendung, wobei wir für die Lösung der auftretenden quadratischen Teilprobleme die Methode der aktiven Menge wählen.
- Die Berechnung des Gradienten und der Hesse-Matrix für die Zielfunktion von $(\widetilde{\text{KQ2}})$ erfolgt in jeder Iteration mittels Satz 5.1.2.

²⁸ $r_{\min} := \min_{w \in \mathbb{R}^n} \{ \mu^\top w \mid I_n w \geq 0, e^\top w = 1 \}, r_{\max} := \max_{w \in \mathbb{R}^n} \{ \mu^\top w \mid I_n w \geq 0, e^\top w = 1 \}.$

5. Portfoliooptimierung in der Praxis

- Wir berechnen zu Beginn mithilfe des konvexen Optimierungsproblems $(LS_{\hat{\theta}}^c)$ das „echte“ „Risk-Parity“-Portfolio w^{RP} mit positiven Gewichten und Gewichtssumme 1, da wir in Abschnitt 5.2.1 gesehen haben, dass in der Praxis bei direkter Verwendung der Zielfunktion f_2 durchaus lokale Optima auftreten können, die nicht global optimal sind.
- Nun starten wir bei $r_{RP} := \mu^T w^{RP}$ und erhöhen zuerst iterativ die Return-Schranke bis r_{max} , anschließend verringern wir diese (wieder ausgehend von r_{RP}) bis r_{min} , sodass wir insgesamt auch für $(\widehat{KQ2})$ 300 Portfolios berechnen. Als Startpunkt wählen wir zu Beginn (w^0, c^0) mit $w^0 = w^{RP}$ und $c^0 = (w^{RP})^T \Sigma w^{RP}$, für die weiteren Berechnungen wieder den jeweils zuletzt berechneten Punkt. Mit dieser Vorgangsweise können wir davon ausgehen, stets einen „guten“ (also in der Nähe des zu berechnenden approximativen „Risk-Parity“-Portfolios liegenden) Startpunkt zu verwenden, da die Return-Schranke pro Berechnung nur geringfügig um $\frac{r_{max} - r_{min}}{300}$ geändert wird. Dieser Startpunkt ist zwar bei der iterativen *Erhöhung* der Return-Schranke nicht zulässig, dies stellt jedoch für unseren SQP-Algorithmus kein Problem dar (vgl. die diesbezügliche Bemerkung in Abschnitt 5.1). Die zuletzt verwendete BFGS-Matrix eines Problems verwenden wir wiederum als Startmatrix für die BFGS-Formel im darauffolgenden Problem mit geringfügig höherer bzw. niedrigerer Return-Schranke. Verschiedene Versuche haben gezeigt, dass dies eine sinnvolle Vorgangsweise ist und bei Verwendung der n -dimensionalen Einheitsmatrix als BFGS-Startmatrix (Defaultwert in unserer Implementierung) oftmals mehr Iterationen benötigt werden.
































































DJIA	DJIA	DJTA	DJUA
 AA	 MMM	 ALK	 AEP
 AXP	 MRK	 CHRW	 AES
 BA	 MSFT	 CNW	 CNP
 BAC	 PFE	 CSX	 D
 CAT	 PG	 EXPD	 DUK
 CSCO	 T	 FDX	 ED
 CVX	 TRV	 GMT	 EIX
 DD	 UNH	 JBHT	 EXC
 DIS	 UTX	 JBLU	 FE
 GE	 VZ	 KEX	 NEE
 HD	 WMT	 KSU	 NI
 HPQ	 XOM	 LSTR	 PCG
 IBM		 LUV	 PEG
 INTC		 MATX	 SO
 JNJ		 NSC	 WMB
 JPM		 R	
 KO		 UNP	
 MCD		 UPS	

Abb. 5.1.: Farbcodes der 63 Aktien aus dem DJCA

5. Portfoliooptimierung in der Praxis

Für sämtliche folgende Grafiken kommt die in Abbildung 5.1 angegebene Legende zur Anwendung. Damit können wir anhand der farblichen Zuordnung auch für den zusammengesetzten *Dow Jones Composite Average* auf einen Blick erkennen, aus welchen Subindizes die einzelnen Aktien stammen.

5.2.2.1. Vergleich der Zeiträume 2004/2005 und 2008/2009

Wir beginnen mit dem 63 Aktien beinhaltenden *Dow Jones Composite Average*. Wie bereits erwähnt, haben wir die Schätzungen für μ und Σ für die folgenden Untersuchungen aus den täglichen Returns von 1. Januar 2004 bis 31. Dezember 2005 ermittelt.

In Abbildung 5.2 sehen wir die Zusammensetzung der berechneten „Mean-Variance“-Portfolios. Dort ist auch das „Minimum-Variance“-Portfolio eingezeichnet, welches im betrachteten Zeitraum 2004/2005 einen erwarteten täglichen Return von 0.0265% aufweist. Abbildung 5.3 zeigt uns dagegen die Zusammensetzung der approximativen „Risk-Parity“-Portfolios. Der erwartete tägliche Return des „echten“ „Risk-Parity“-Portfolios liegt bei 0.0526% und damit etwa doppelt so hoch wie für „Minimum-Variance“, dessen Returnwert zum Vergleich ebenfalls in der Grafik aufscheint.

Uns fällt sofort auf, dass die Gewichte im „Risk-Parity“-Portfolio recht gleichmäßig über alle Aktien verteilt sind. Jede Aktie ist positiv gewichtet. Wie wir aus den theoretischen Untersuchungen in Kapitel 3.3.1 bereits wissen, muss dies auch so sein (Lemma 3.3.1). Für unsere approximativen Lösungen mit höheren Return-Schranken werden zwar auch Aktien mit 0 gewichtet, trotzdem erhalten wir im Verhältnis zu „Mean-Variance“ großteils eine viel stärkere Diversifikation. Das „Minimum-Variance“-Portfolio setzt sich beispielsweise aus nur 19 Aktien (davon nur eine aus dem DJTA) zusammen. Vor allem ist es auch gelungen, die Return-Restriktion $\mu^\top w \geq r$ vom „Mean-Variance“- auf das „Risk-Parity“-Modell zu übertragen, womit einem Investor ermöglicht wird, Anforderungen an den Mindest-Return zu stellen und trotzdem die Idee des „Risk-Parity“ zu nutzen. Für die verwendete Testinstanz scheint somit das Optimierungsproblem ($\widetilde{\text{KQ2}}$) seinen Zweck zu erfüllen.

In den nachfolgenden Grafiken sind die analogen Ergebnisse für die Teilindizes DJIA (Abbildungen 5.4 und 5.5), DJTA (Abbildungen 5.6 und 5.7) sowie DJUA (Abbildungen 5.8 und 5.9) angegeben. Wir können dort wieder die starke Diversifikation bei „Risk-Parity“ bzw. eine Konzentration auf wenige Aktien bei „Mean-Variance“ erkennen. Wie beim DJCA liegt der Return des „echten“ „Risk-Parity“-Portfolios jeweils über jenem der „Minimum-Variance“-Lösung. Auch für höhere Return-Schranken scheint es, als würden wir sinnvolle Portfolios basierend auf einer Approximation des „Risk-Parity“-Zustandes erhalten.

Abbildung 5.10 beinhaltet nun die interessante Gegenüberstellung der durchschnittlichen realisierten *täglichen* Portfolio>Returns und der entsprechenden realisierten Portfolio-Varianzen der „Mean-Variance“- und der approximativen „Risk-Parity“-Portfolios, welche wir für den zusammengesetzten DJCA-Index ermittelt haben. Wir vergleichen diese Werte sowohl für die Schätzperiode 2004/2005 als auch für den Zeitraum 2008/2009 (also während der

5. Portfoliooptimierung in der Praxis

Finanzkrise). In den oberen beiden Plots (Schätzperiode 2004/2005) ist auf der x -Achse die untere Schranke für den Portfolio-Return angegeben, die wir bei der Berechnung der Portfolios verwendet haben. Wie bereits erläutert, wurden 300 Portfolios mit äquidistanten Werten im Intervall $[r_{\min}, r_{\max}]$ für die untere Return-Schranke berechnet. Auf den unteren beiden Plots (Vergleichszeitraum 2008/2009) gibt die x -Achse hingegen den korrespondierenden Index dieser 300 Portfolios an. Der Index 0 entspricht dabei jenem Portfolio, welches zuvor unter Verwendung der unteren Schranke r_{\min} ermittelt wurde, während der Index 300 jenes Portfolio angibt, dessen Konstruktion mithilfe der unteren Schranke r_{\max} erfolgte. Man beachte, dass die Grafik rechts oben die effiziente Front beinhaltet (vgl. Kapitel 2). Diese entspricht genau der blauen Linie für alle Return-Schranken „rechts“ der „Minimum-Variance“-Lösung (blauer Punkt). Des Weiteren beobachten wir wie folgt:

- **Zeitraum 01.01.2004 bis 31.12.2005:**

Die Varianzen der approximativen „Risk-Parity“-Lösungen liegen nur minimal über jenen der „Mean-Variance“-Portfolios. Dies ist dahingehend bemerkenswert, dass bei $(\widetilde{KQ2})$ nur versucht wird, den „Risk-Parity“-Zustand zu erreichen, während in $(\widetilde{MW1})$ eine Varianzminimierung stattfindet. Ein potentieller Nachteil des „Risk-Parity“-Modells gegenüber dem Markowitz’schen Modell (nämlich höhere Volatilitäten) scheint hier kein allzu großes Problem darzustellen.

- **Zeitraum 01.01.2008 bis 31.12.2009:**

Wir machen in der Grafik links unten (Abb. 5.10) im Wesentlichen folgende interessante Beobachtung: Je *höher* wir zuvor für 2004/2005 die untere Schranke für den erwarteten täglichen Portfolio-Return gewählt haben, desto *niedriger* fällt schlussendlich der realisierte tägliche Return im Zeitraum 2008/2009 aus. Über weite Bereiche liefern jedoch alle 300 berechneten approximativen „Risk-Parity“-Portfolios höhere Returns als die korrespondierenden „Mean-Variance“-Lösungen und scheinen damit krisenfester zu sein. Für die Portfolio-Indizes 160 bis 225 (Return-Schranken zwischen 0.074% und 0.130% Prozent bei der Berechnung der Portfolios anhand der Daten von 1. Januar 2004 bis 31. Dezember 2005) erhalten wir beispielsweise bei „Risk-Parity“ noch positive Returns, bei „Mean-Variance“ dagegen negative. Zudem liegen für die Portfolios mit den Indizes von 150 bis 300 (was zuvor einer unteren Return-Schranke zwischen 0.065% und $r_{\max} \approx 0.194\%$ entsprochen hat) die realisierten Varianzen auf nahezu identischem Niveau. Wir würden daher a posteriori wohl die approximativen „Risk-Parity“-Portfolios bevorzugen.

5. Portfoliooptimierung in der Praxis

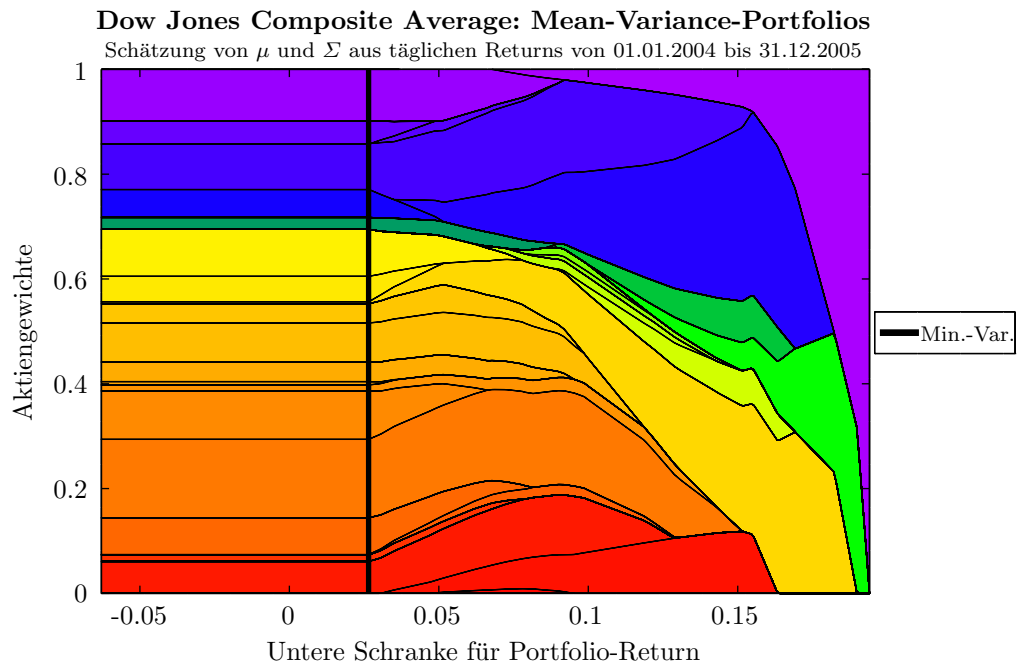


Abb. 5.2.: DJCA: „Mean-Variance“-Portfolios des Zeitraumes 2004/2005

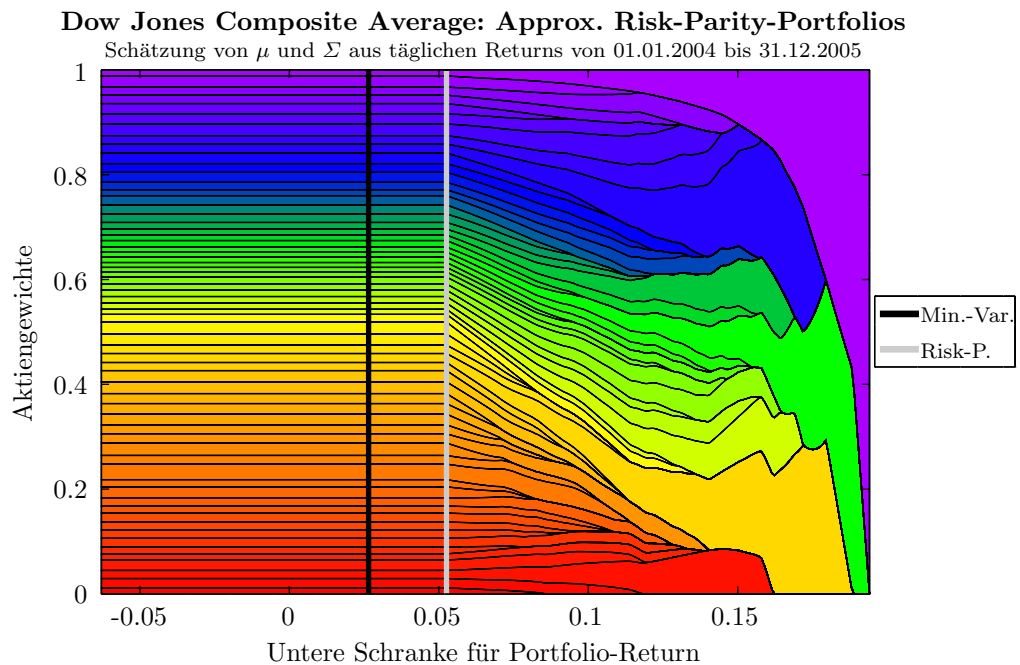


Abb. 5.3.: DJCA: approximative „Risk-Parity“-Portfolios des Zeitraumes 2004/2005

5. Portfoliooptimierung in der Praxis

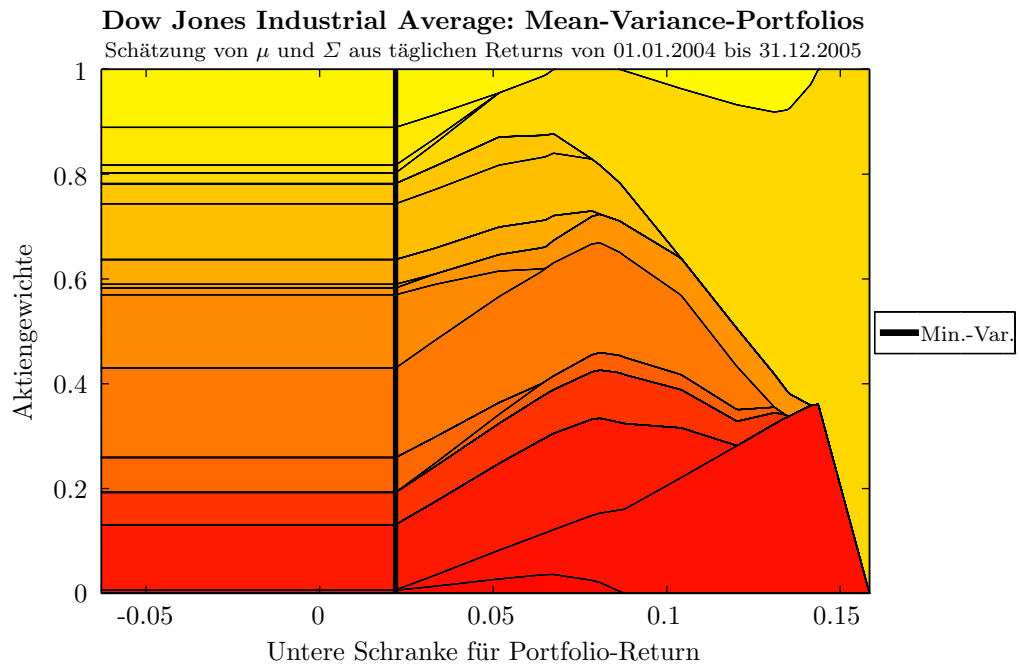


Abb. 5.4.: DJIA: „Mean-Variance“-Portfolios des Zeitraumes 2004/2005

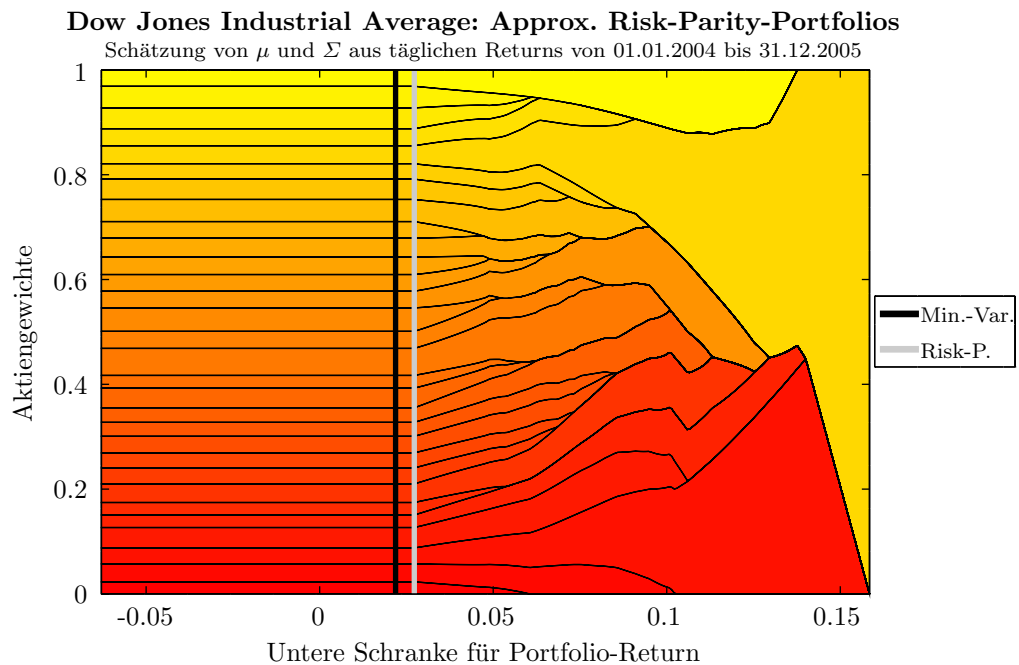


Abb. 5.5.: DJIA: approximative „Risk-Parity“-Portfolios des Zeitraumes 2004/2005

5. Portfoliooptimierung in der Praxis

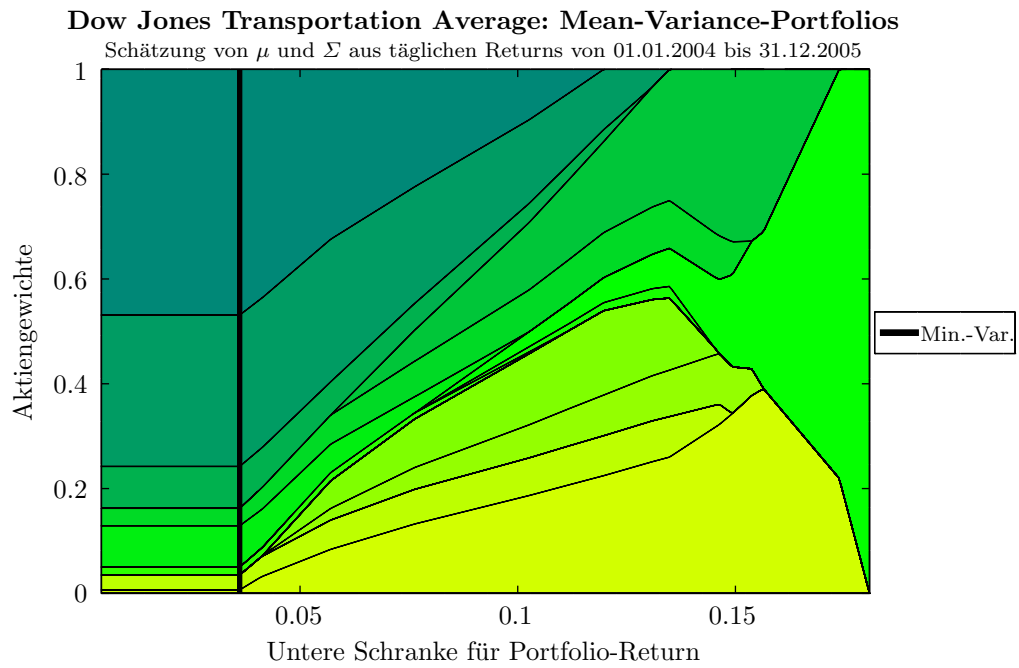


Abb. 5.6.: DJTA: „Mean-Variance“-Portfolios des Zeitraumes 2004/2005

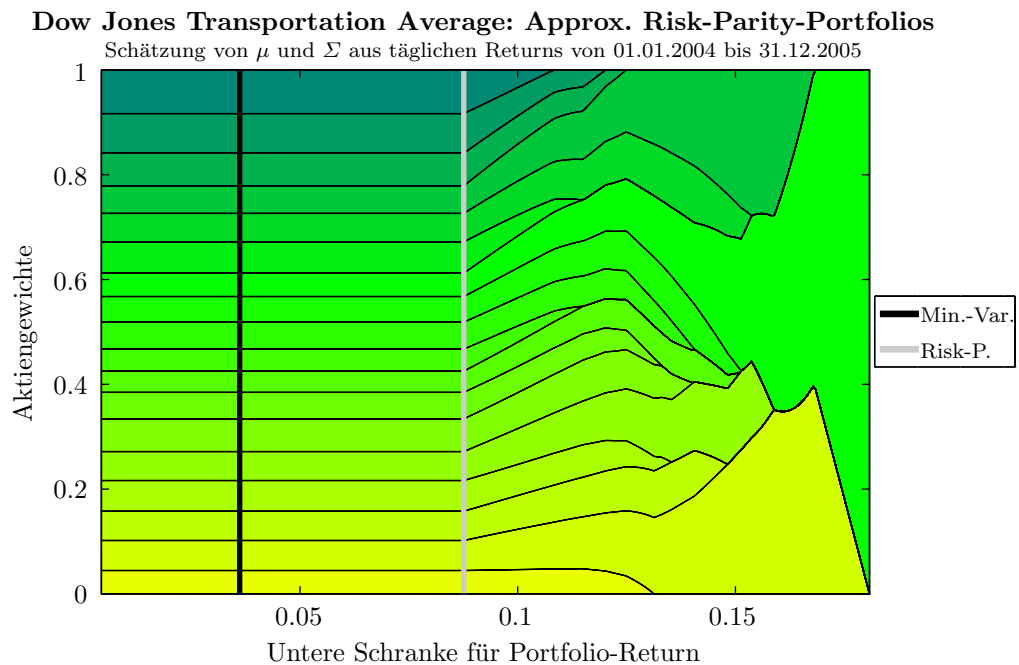


Abb. 5.7.: DJTA: approximative „Risk-Parity“-Portfolios des Zeitraumes 2004/2005

5. Portfoliooptimierung in der Praxis

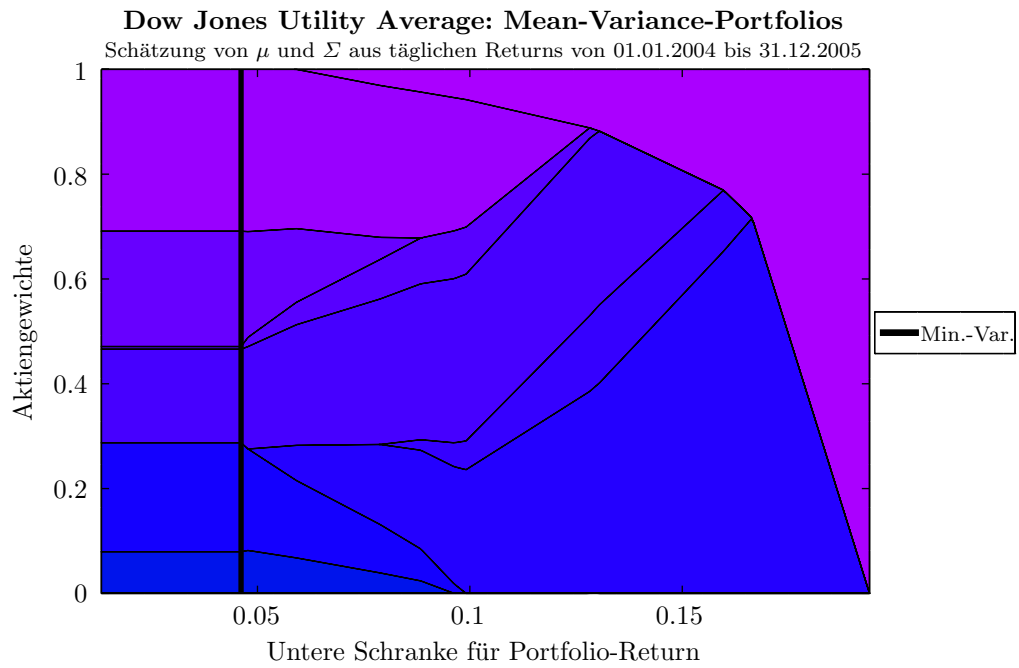


Abb. 5.8.: DJUA: „Mean-Variance“-Portfolios des Zeitraumes 2004/2005

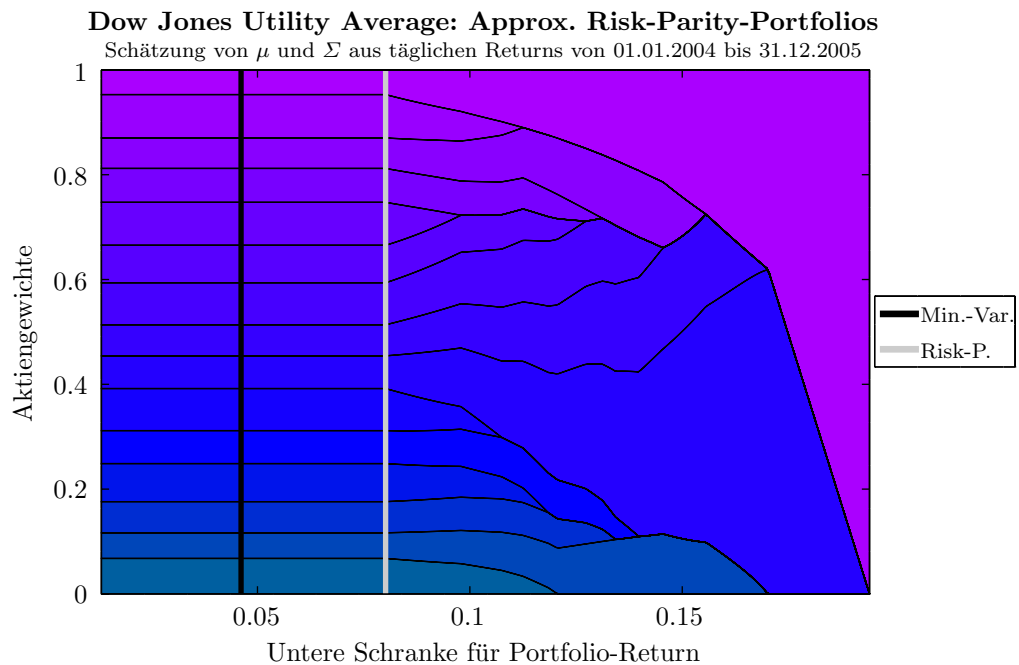


Abb. 5.9.: DJUA: approximative „Risk-Parity“-Portfolios des Zeitraumes 2004/2005

5. Portfoliooptimierung in der Praxis

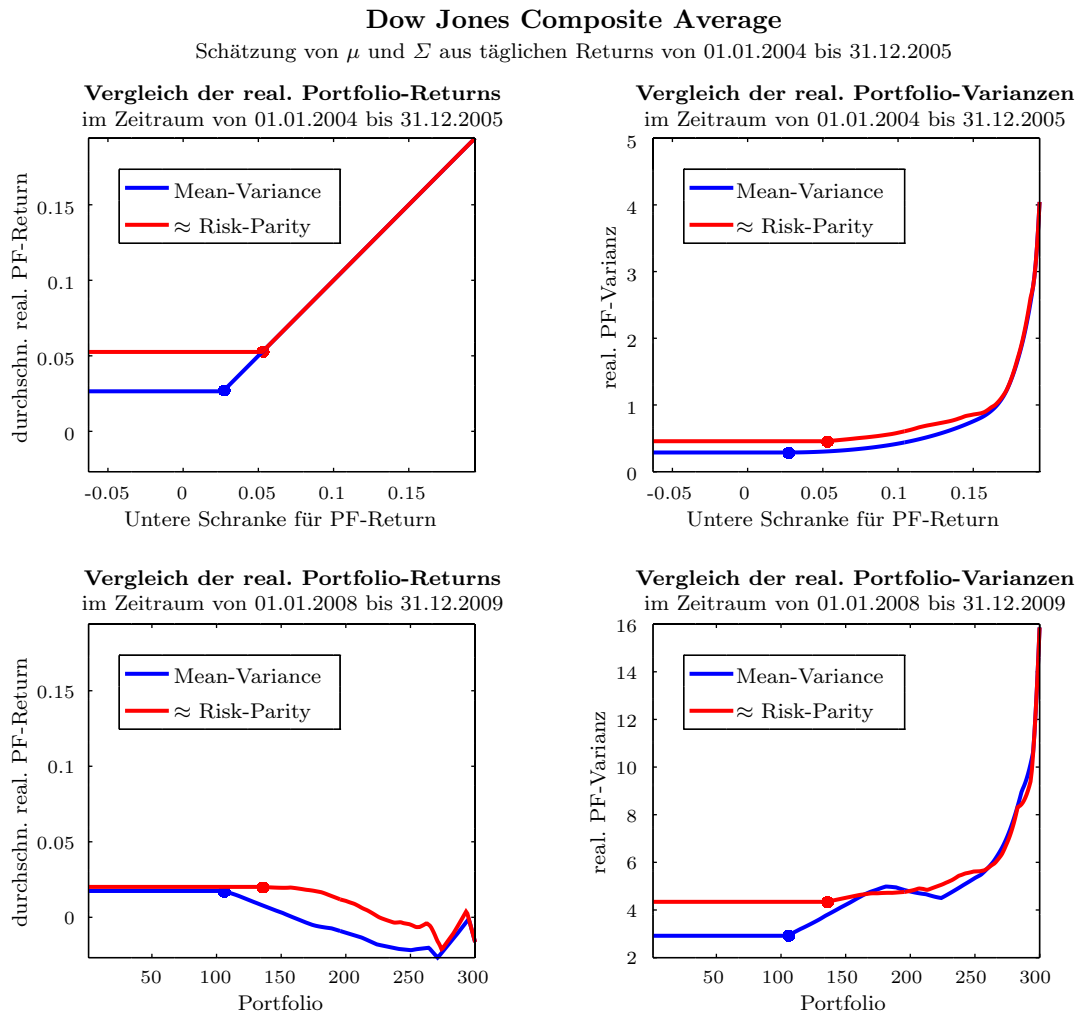


Abb. 5.10.: DJCA: „Mean-Variance“- und approx. „Risk-Parity“-Portfolios von 2004/2005 im Vergleich
blauer Punkt: „Minimum-Variance“-Portfolio
roter Punkt: „Risk-Parity“-Portfolio

5. Portfoliooptimierung in der Praxis

5.2.2.2. Vergleich der Zeiträume 2008/2009 und 2013/2014

Verwenden wir nun die Zeitspanne von 1. Januar 2008 bis 31. Dezember 2009 für die Schätzung von Σ und μ , so erkennen wir in den Abbildungen 5.11 bis 5.18 bei den „Mean-Variance“-Portfolios vor allem für die Indizes DJCA, DJIA und DJUA eine noch extremere Konzentration auf wenige Aktien. Während im vorhergehenden Kapitel (Schätzung aus dem Zeitraum 2004/2005) das „Minimum-Variance“-Portfolio für den DJCA noch 19 positiv gewichtete Aktien beinhaltet hat, sind es hier nur noch 7 (von 63) Aktien (siehe Abb. 5.11). Die „Minimum-Variance“-Lösung bei Verwendung des DJUA, welcher aus 15 Aktien besteht, beinhaltet gar nur 2 Aktien (vgl. Abb. 5.17). Eine starke Risikokonzentration ist die Folge. Dagegen weisen die „Risk-Parity“-Lösungen der beiden betrachteten Zeiträume optisch keine so großen Unterschiede auf (vgl. Abb. 5.3 und 5.12). Dies lässt vermuten, dass das „Mean-Variance“-Modell heftiger auf Änderungen der Inputwerte reagiert als unser „Risk-Parity“-Modell. Eine prinzipiell hohe Sensitivität des Markowitz’schen Ansatzes bezüglich der Input-Parameter wurde unter anderem auch von Merton (1980) sowie Best und Grauer (1991) beobachtet. Interessanterweise liegt der erwartete Return des „Risk-Parity“-Portfolios für den DJCA sowie die zwei Teilindizes DJIA und DJUA unter jenem von „Minimum-Variance“. Nur für den DJTA ist dies umgekehrt, so wie wir es im ersten Zeitraum 2004/2005 für alle vier Indizes beobachtet haben.

Wir interpretieren nun abschließend noch Abbildung 5.19 mit der Gegenüberstellung von „Mean-Variance“ und „Risk-Parity“ für den *Dow Jones Composite Average*, welche analog zu Abbildung 5.10 aufgebaut ist: Die beiden Grafiken in der ersten Zeile behandeln wieder den Zeitraum der Schätzung (hier also 2008/2009), die unteren beiden Grafiken zeigen dagegen die durchschnittlichen realisierten täglichen Returns und die realisierten Varianzen der von uns berechneten Portfolios fünf Jahre später (2013/2014):

- **Zeitraum 01.01.2008 bis 31.12.2009:**

In der ersten Grafik sehen wir nochmals, dass der Return der „Risk-Parity“-Lösung unter jenem des „Mean-Variance“-Portfolios liegt. In der zweiten Grafik liegt die rote Linie mit den Varianzen der approximativen „Risk-Parity“-Portfolios etwas höher über der effizienten Front (blau) als dies in Abbildung 5.10 des vorhergehenden Abschnittes der Fall war. Dies ist vermutlich darauf zurückzuführen, dass sich im hier betrachteten Zeitraum während der Krise die Volatilitäten der Returns für die einzelnen Aktien stärker voneinander unterscheiden (vgl. die Tabelle am Beginn von Kapitel 5.2.2). Trotzdem liegen die beiden Linien nicht extrem weit auseinander.

- **Zeitraum 01.01.2013 bis 31.12.2014:**

Die approximativen „Risk-Parity“-Portfolios liefern aus der Sicht eines Investors wohl deutlich bessere Ergebnisse. Diese weisen in der Periode 2013/2014 teilweise deutlich höhere realisierte Returns auf und für die Portfolio-Indizes 176 bis 300 (was bei der Portfolioberechnung einer Return-Schranke von r zwischen 0.078% und 0.203% entsprach) sogar auch etwas kleinere Varianzen als die „Mean-Variance“-Äquivalente. Beim Vergleich der Skalen der beiden Grafiken mit den Portfolio-Varianzen erkennen

5. Portfoliooptimierung in der Praxis

wir jedoch auch, dass diese vom Zeitraum 2008/2009 auf den Zeitraum 2013/2014 generell wieder auf ein viel niedrigeres Niveau sinken.

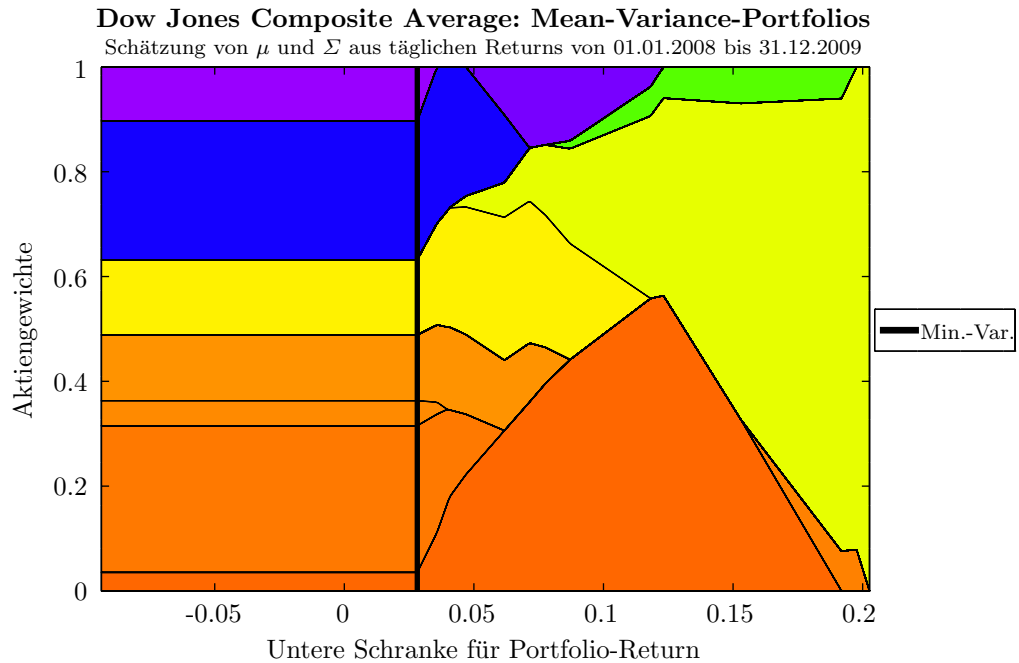


Abb. 5.11.: DJCA: „Mean-Variance“-Portfolios des Zeitraumes 2008/2009

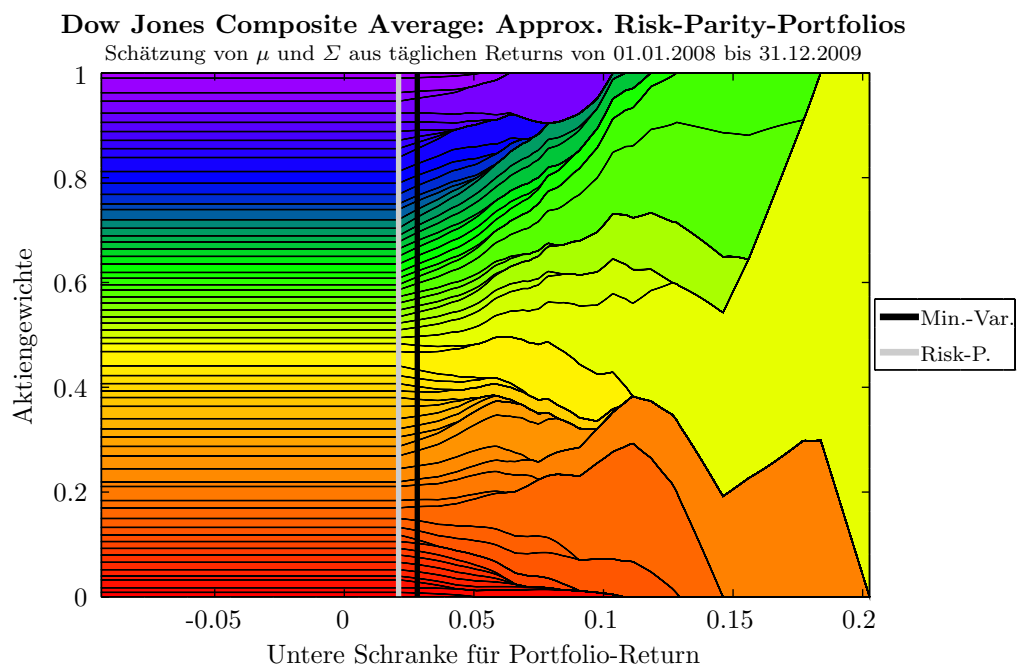


Abb. 5.12.: DJCA: approximative „Risk-Parity“-Portfolios des Zeitraumes 2008/2009

5. Portfoliooptimierung in der Praxis

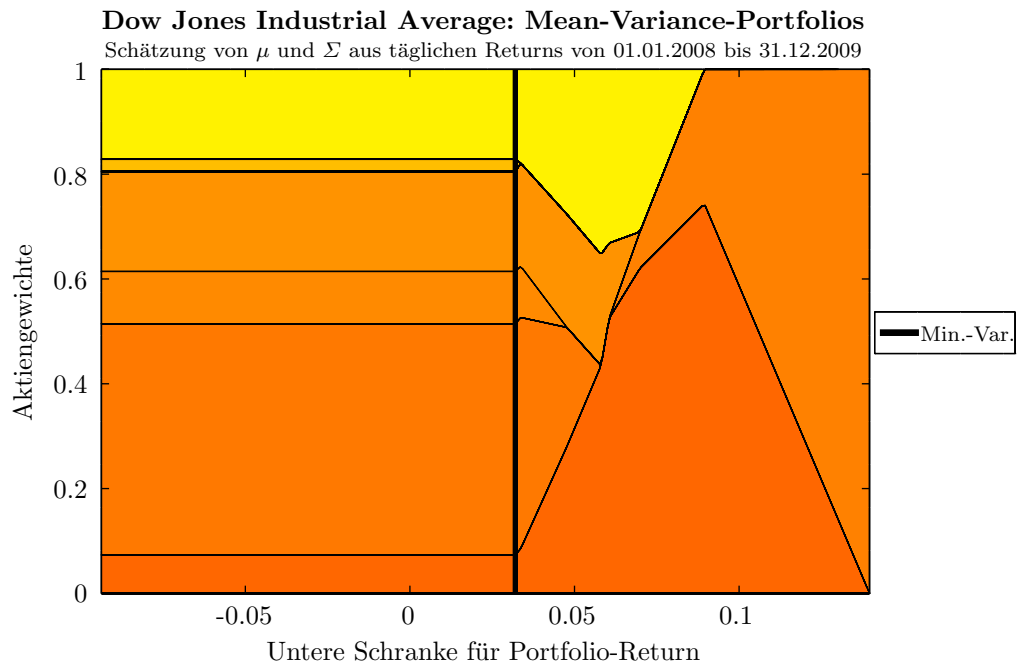


Abb. 5.13.: DJIA: „Mean-Variance“-Portfolios des Zeitraumes 2008/2009

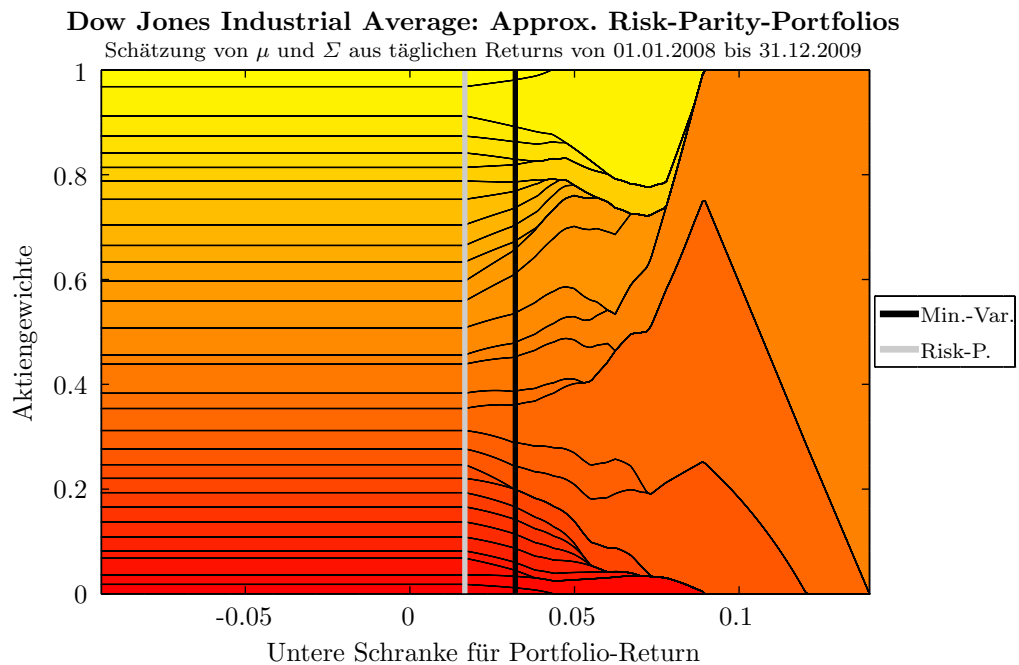


Abb. 5.14.: DJIA: approximative „Risk-Parity“-Portfolios des Zeitraumes 2008/2009

5. Portfoliooptimierung in der Praxis

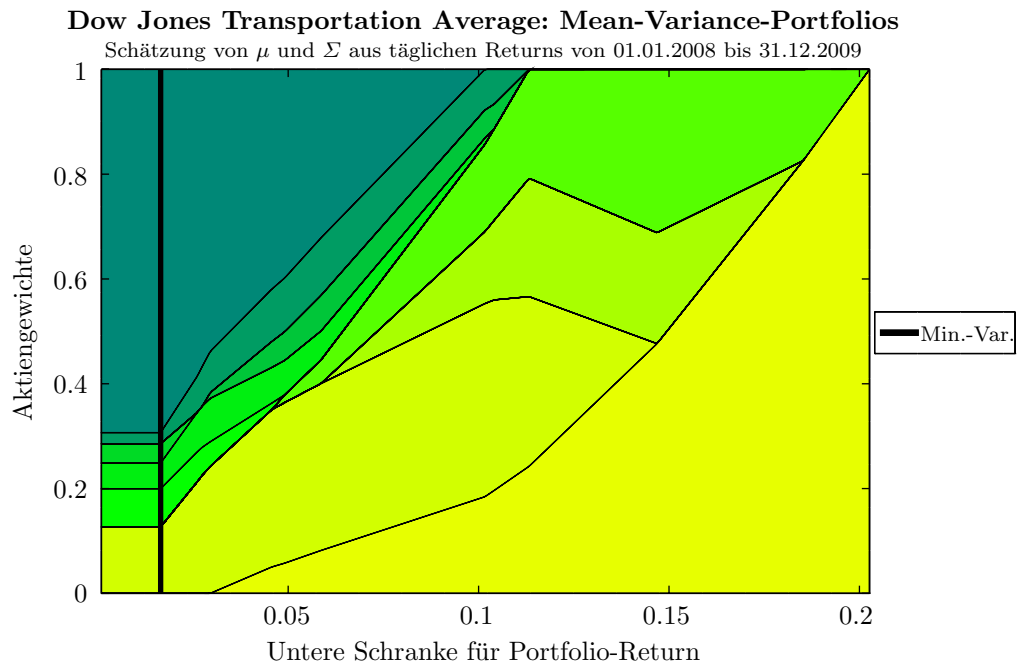


Abb. 5.15.: DJTA: „Mean-Variance“-Portfolios des Zeitraumes 2008/2009

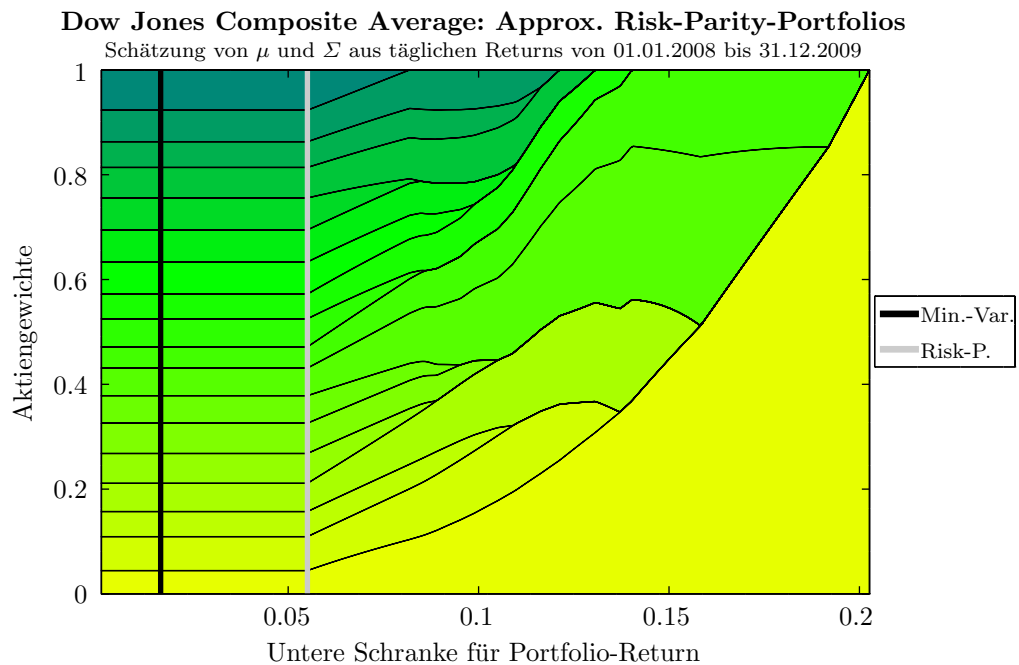


Abb. 5.16.: DJTA: approximative „Risk-Parity“-Portfolios des Zeitraumes 2008/2009

5. Portfoliooptimierung in der Praxis

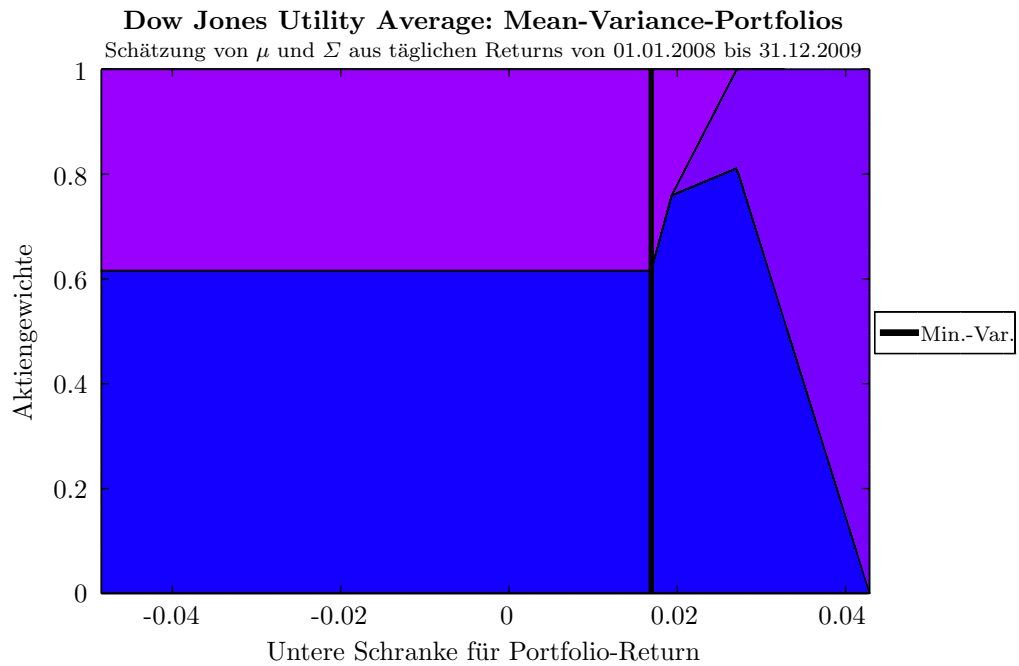


Abb. 5.17.: DJUA: „Mean-Variance“-Portfolios des Zeitraumes 2008/2009

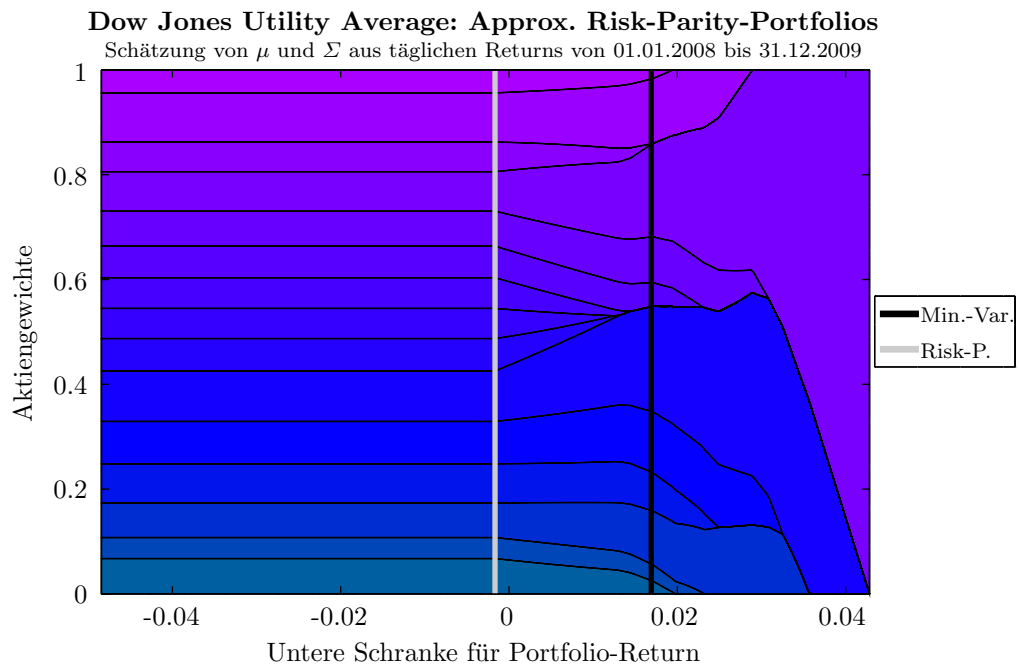


Abb. 5.18.: DJUA: approximative „Risk-Parity“-Portfolios des Zeitraumes 2008/2009

5. Portfoliooptimierung in der Praxis

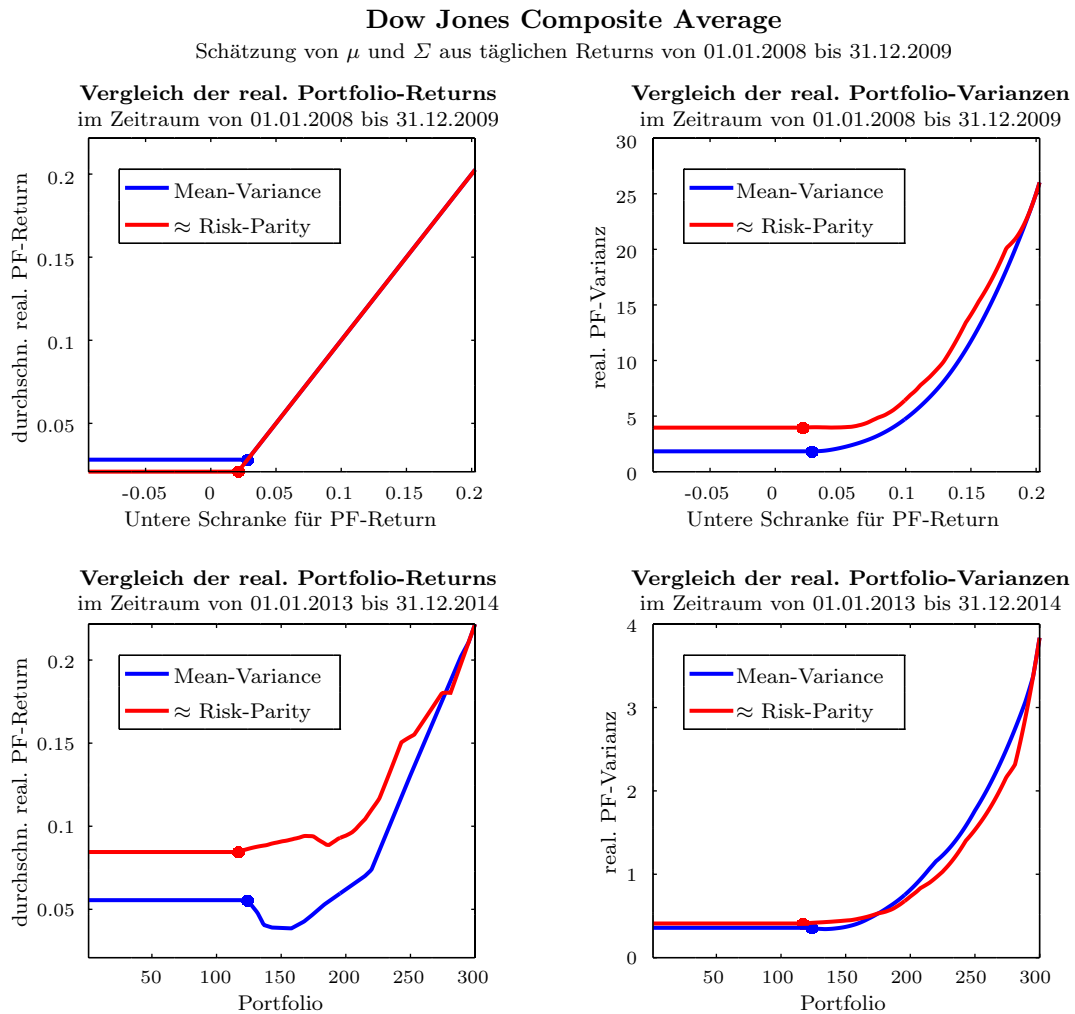


Abb. 5.19.: DJCA: „Mean-Variance“- und approx. „Risk-Parity“-Portfolios von 2008/2009 im Vergleich
blauer Punkt: „Minimum-Variance“-Portfolio
roter Punkt: „Risk-Parity“-Portfolio

6. Conclusio und Ausblick

Im theoretischen Teil dieser Arbeit wurde das grundlegende Markowitz'sche Modell definiert und ausgehend vom „Minimum-Variance“-Modell die risikobasierte Portfoliooptimierung eingeführt. Für das „Risk-Budgeting“-Modell war es uns möglich, anhand eines konvexen Optimierungsproblems mit logarithmischer Schranke interessante Aspekte dieses Ansatzes zu analysieren und Kleinste-Quadrate-Formulierungen anzugeben, deren globale Optima unter gewissen Voraussetzungen „Risk-Budgeting“-Portfolios sind. Anschließend haben wir verschiedene Methoden zur Lösung nichtlinearer Programme kennengelernt, wobei bei der Auswahl der vorgestellten Verfahren die Struktur der erwähnten Kleinste-Quadrate-Optimierungsprobleme ausschlaggebend war.

Im praktischen Teil haben wir diese Optimierungsalgorithmen implementiert und festgestellt, dass damit die betrachteten Optimierungsprobleme im Zusammenhang mit dem „Mean-Variance“- und dem „Risk-Budgeting“-Modell prinzipiell behandelbar sind. Wir konnten jedoch auch erkennen, dass die Nicht-Konvexität der Kleinste-Quadrate-Formulierungen des „Risk-Budgeting“-Ansatzes unter Umständen Probleme bereiten kann. Daher war auch die betrachtete Formulierung als konvexes Programm mit logarithmischer Schranke nützlich.

Ein Vergleich des Markowitz'schen und des „Risk-Parity“-Ansatzes (als Spezialfall von „Risk-Budgeting“) anhand realer Finanzdaten hat vielversprechende Ergebnisse geliefert. Die approximativen „Risk-Parity“-Portfolios waren auf der betrachteten Aktienmenge stark diversifiziert und können damit generell als robuster gegenüber unerwartet niedrigen Returns oder hohen Varianzen einzelner Aktien angesehen werden. Eine vergleichende Analyse von durchschnittlichen realisierten täglichen Returns und realisierten Varianzen der berechneten Portfolios für verschiedene Zeiträume im Umfeld der Finanzkrise, die Mitte 2007 begann, hat eine gute Performance von „Risk-Parity“ gezeigt. So resultieren beispielsweise für die approximativen „Risk-Parity“-Portfolios, die wir für den *Dow Jones Composite Average* Aktienindex anhand der Aktienkurse aus den Jahren 2004 und 2005 berechnet haben, vier Jahre später (also während der Finanzkrise) teils deutlich höhere Returns wie für die korrespondierenden „Mean-Variance“-Portfolios, während die realisierten Varianzen auf nahezu identischem Niveau liegen.

Weiterführende Untersuchungen im Bereich des „Risk-Budgeting“ sind beispielsweise im Bereich der Verwendung anderer Risikomaße anstatt der Varianz denkbar. Auch die Anwendung anderer Methoden für die Lösung der besprochenen Kleinste-Quadrate-Probleme bietet noch weitreichende Forschungsmöglichkeiten. Bai, Scheinberg und Tütüncü (2013)

6. Conclusio und Ausblick

haben diesbezüglich beispielsweise eine „Alternating Linearization“-Methode (ALM) betrachtet und gute Resultate erzielt. Ebenso wären noch Recherchen dahingehend denkbar, das „Risk-Budgeting“-Problem in Form eines gänzlich anderen Optimierungsproblems zu behandeln. Dabei könnte von besonderem Interesse sein, den Vorteil der Konvexität unseres Programms mit logarithmischer Schranke und die Möglichkeit der nicht-konvexen Kleinste-Quadrate-Probleme, beliebige Gleichheits- und Ungleichheitsrestriktionen einzubauen, zu kombinieren.

Appendix

Anhang A.

Appendix

A.1. Pareto-Optimalität

(Siehe Brito und Vicente, 2012.)

Betrachte folgendes nichtlineare, multikriterielle Optimierungsproblem:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & F(x) \equiv (f_1(x), \dots, f_m(x))^\top \\ \text{s.t.} \quad & x \in \Omega \subseteq \mathbb{R}^n \end{aligned}$$

mit m Zielfunktionen $f_i : \mathbb{R}^n \rightarrow \mathbb{R}, \forall i = 1, \dots, m$.

Das Konzept der Pareto-Optimalität liefert eine Möglichkeit, zwei verschiedene Punkte $x \in \Omega$ bezüglich ihrer Optimalität für multikriterielle Probleme miteinander zu vergleichen:

Definition. Seien $x, y \in \Omega$ zwei zulässige Punkte für obiges multikriterielle Optimierungsproblem. Dann sagen wir „ x **dominiert** y “, genau dann wenn

$$F(y) - F(x) \in \mathbb{R}_+^m \setminus \{0\}.$$

Diese Definition ist äquivalent zur Forderung, dass $f_i(x) \leq f_i(y), \forall i = 1, \dots, m$ und $\exists j \in \{1, \dots, m\}$ mit $f_j(x) < f_j(y)$.

Definition. Ein zulässiger Punkt $x^* \in \Omega$ heißt **Pareto-optimal** (oder Pareto-effizient) für F auf Ω , wenn kein $y \in \Omega$ existiert, sodass y den Punkt x^* dominiert.

A.2. Positiv homogene Funktionen

(Siehe Bai, Scheinberg und Tütüncü, 2013.)

Definition. Eine Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ heißt **positiv homogen vom Grad k** , falls für eine beliebige Konstante $c \in \mathbb{R}^+$ und einen beliebigen Punkt $x \in \mathbb{R}^n$

$$f(c \cdot x) = c^k \cdot f(x)$$

gilt.

Satz A.2.1. (Satz von Euler für positiv homogene Funktionen)

Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ eine positiv homogene Funktion vom Grad k . Dann gilt:

$$k \cdot f(x) = \sum_{i=1}^n x_i \cdot \frac{\partial f(x)}{\partial x_i}$$

Beweis. Es gelten die folgenden beiden Gleichungen:

$$\frac{\partial f(c \cdot x)}{\partial c} = \frac{\partial (c^k \cdot f(x))}{\partial c} = k \cdot c^{k-1} \cdot f(x)$$

$$\frac{\partial f(c \cdot x)}{\partial c} = \sum_{i=1}^n \frac{\partial f(c \cdot x)}{\partial x_i} \cdot \frac{\partial x_i}{\partial (c \cdot x_i)} \cdot \frac{\partial (c \cdot x_i)}{\partial c} = \sum_{i=1}^n c^k \cdot \frac{\partial f(x)}{\partial x_i} \cdot \frac{1}{c} \cdot x_i = c^{k-1} \sum_{i=1}^n \frac{\partial f(x)}{\partial x_i} \cdot x_i$$

Damit folgt sofort die zu beweisende Identität. □

A.3. Eigenschaften der konstanten Korrelationsmatrix

(Siehe Maillard, Roncalli und Teiletche (2010) sowie Roncalli (2013).)

Sei

$$P = P_n(\varrho) := \begin{pmatrix} 1 & \varrho & \cdots & \varrho \\ \varrho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \varrho \\ \varrho & \cdots & \varrho & 1 \end{pmatrix}$$

eine positiv definite $n \times n$ -Korrelationsmatrix mit konstanten Korrelationen. Wir möchten mithilfe der komplementären Matrix die Inverse von P herleiten. Weiters werden wir zeigen, dass P genau dann positiv semidefinit und damit eine zulässige Korrelationsmatrix ist, falls ϱ im abgeschlossenen Intervall $[-\frac{1}{n-1}, 1]$ liegt, und genau dann positiv definit ist, falls ϱ einen Wert aus dem entsprechenden offenen Intervall annimmt.

Anhang A. Appendix

Wir beginnen mit der Berechnung der Determinante, wofür wir folgende Zeilenumformungen durchführen:

$$\begin{aligned}
 |P| &= \begin{vmatrix} 1 & \varrho & \varrho & \cdots & \cdots & \varrho \\ 0 & 1-\varrho & \varrho-1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & 1-\varrho & \varrho-1 & 0 \\ 0 & \cdots & \cdots & 0 & 1-\varrho & \varrho-1 \\ \varrho & \varrho & \cdots & \cdots & \varrho & 1 \end{vmatrix} = \begin{vmatrix} 1 & \varrho & \varrho & \cdots & \cdots & \varrho \\ 0 & 1-\varrho & \varrho-1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & 1-\varrho & \varrho-1 & 0 \\ 0 & \cdots & \cdots & 0 & 1-\varrho & \varrho-1 \\ 0 & \varrho-\varrho^2 & \cdots & \cdots & \varrho-\varrho^2 & 1-\varrho^2 \end{vmatrix} \\
 &= \begin{vmatrix} 1 & \varrho & \varrho & \cdots & \cdots & \varrho \\ 0 & 1-\varrho & \varrho-1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & 1-\varrho & \varrho-1 & 0 \\ 0 & \cdots & \cdots & 0 & 1-\varrho & \varrho-1 \\ 0 & 0 & \cdots & 0 & 0 & (1-\varrho^2) + (n-2)(\varrho-\varrho^2) \end{vmatrix}
 \end{aligned}$$

Damit folgt

$$|P| = (1-\varrho)^{n-2} \cdot [(1-\varrho^2) + (n-2)(\varrho-\varrho^2)] = (1-\varrho)^{n-2} \cdot [1 + (n-2)\varrho - (n-1)\varrho^2]$$

Sei $\tilde{P} = (\tilde{p}_{i,j})_{1 \leq i,j \leq n}$ die komplementäre Matrix von P . Dann gilt unter Beachtung der Tatsache, dass sowohl P als auch \tilde{P} symmetrisch sind,

$$\tilde{p}_{i,j} = (-1)^{i+j} |P^{i,j}|,$$

wobei $P^{i,j}$ jene Matrix ist, welche aus P durch Streichung der i -ten Zeile und j -ten Spalte entsteht. Für die Diagonale erhalten wir offensichtlich $\tilde{p}_{i,i} = |P_{n-1}(\varrho)|$. Diese Determinante, welche wir bereits oben berechnet haben, lässt sich folgendermaßen umschreiben:

$$\begin{aligned}
 |P_{n-1}(\varrho)| &= (1-\varrho)^{n-3} \cdot [(1-\varrho^2) + (n-3)(\varrho-\varrho^2)] \\
 &= (1-\varrho)^{n-2} \cdot \left[\frac{(1-\varrho)(1+\varrho) + (n-3)\varrho(1-\varrho)}{1-\varrho} \right] \\
 &= (1-\varrho)^{n-2} \cdot [1 + (n-2)\varrho]
 \end{aligned}$$

Für $i < j$ sind alle Einträge der i -ten Spalte und $(j-1)$ -ten Zeile von $P^{i,j}$ gleich ϱ . Daher lässt sich $P^{i,j}$ durch $i-1$ Spalten- und $(n-1)-(j-1) = n-j$ Zeilenvertauschungen in die $(n-1) \times (n-1)$ Matrix

$$\begin{pmatrix} \varrho & 1 & \varrho & \cdots & \varrho \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \varrho \\ \vdots & \ddots & \ddots & \ddots & 1 \\ \varrho & \cdots & \cdots & \cdots & \varrho \end{pmatrix}$$

Anhang A. Appendix

umformen mit einfach zu berechnender Determinante $\varrho(\varrho - 1)^{n-2}$, womit wir

$$\tilde{p}_{i,j} = (-1)^{i+j}(-1)^{i-1+n-j}\varrho(\varrho - 1)^{n-2} = (-1)^{n-1}\varrho(\varrho - 1)^{n-2} = -\varrho(1 - \varrho)^{n-2}$$

erhalten. Insgesamt gilt

$$P^{-1} = \frac{1}{|P|} \tilde{P} = \frac{(1 - \varrho)^{n-2} \cdot [(1 + (n-2)\varrho + \varrho)I_n - \varrho \cdot ee^\top]}{(1 - \varrho)^{n-2} \cdot [1 + (n-2)\varrho - (n-1)\varrho^2]} = \frac{(1 + (n-1)\varrho)I_n - \varrho \cdot ee^\top}{1 + (n-2)\varrho - (n-1)\varrho^2}.$$

Aus den obigen Berechnungen können wir des Weiteren Schranken für ϱ ableiten:

Die Matrix P ist singulär, wenn $(1 - \varrho^2) + (n-2)(\varrho - \varrho^2) = 0$. Dies ist für $\varrho = 1$ der Fall. Falls $\varrho < 1$, so gilt:

$$(1 - \varrho^2) + (n-2)(\varrho - \varrho^2) = 0 \Leftrightarrow (1 + \varrho) + (n-2)\varrho = 0 \Leftrightarrow \varrho = -\frac{1}{n-1}$$

Wir erhalten also eine zulässige (d.h. positiv semidefinite) Korrelationsmatrix, falls $\varrho \in [-\frac{1}{n-1}, 1]$, wobei die Matrix für $\varrho \in (-\frac{1}{n-1}, 1)$ sogar positiv definit ist. Falls $\varrho < -\frac{1}{n-1}$, folgt beispielsweise für $e = (1, \dots, 1)^\top \in \mathbb{R}^n$, dass

$$e^\top P e = n + n(n-1)\varrho < 0,$$

womit P keine Korrelationsmatrix mehr darstellt.

A.4. Cauchy-Schwarz'sche Ungleichung

Satz A.4.1. (Cauchy-Schwarz'sche Ungleichung für \mathbb{R}^n und das Standard-Skalarprodukt)
Seien $x, y \in \mathbb{R}^n$. Dann gilt

$$\left(\sum_{i=1}^n x_i \cdot y_i \right)^2 \leq \left(\sum_{i=1}^n x_i^2 \right) \left(\sum_{i=1}^n y_i^2 \right).$$

Die Ungleichung ist für $x, y \in \mathbb{R}^n \setminus \{0\}$ genau dann mit Gleichheit erfüllt, wenn x und y linear abhängig sind.

Beweis. Siehe Werner (2011). □

A.5. Abstiegsverfahren

(Siehe Geiger und Kanzow, 1999.)

Möchte man ein (lokales) Minimum einer (stetig differenzierbaren) Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$; $x \mapsto f(x)$ ohne weitere Restriktionen aufspüren, so bietet sich die Idee allgemeiner Abstiegsverfahren an. Diese basieren darauf, mithilfe von *Abstiegsrichtungen* eine iterative Verkleinerung des Funktionswertes zu erzielen:

Definition. (Abstiegsrichtung)

Es seien eine Abbildung $f : \mathbb{R}^n \rightarrow \mathbb{R}$ sowie ein Punkt $x^* \in \mathbb{R}^n$ gegeben. Dann heißt $d \in \mathbb{R}^n$ *Abstiegsrichtung* von f in x^* , wenn ein $\hat{t} > 0$ existiert, sodass

$$f(x^* + td) < f(x^*)$$

für alle $t \in (0, \hat{t}]$ gilt.

Lemma A.5.1. *Es seien eine stetig differenzierbare Abbildung $f : \mathbb{R}^n \rightarrow \mathbb{R}$ sowie ein Punkt $x^* \in \mathbb{R}^n$ gegeben. Falls $\nabla f(x^*)^\top d < 0$ für einen Vektor $d \in \mathbb{R}^n$ gilt, so ist d eine Abstiegsrichtung von f in x^* .*

Beweis. (Vgl. Geiger und Kanzow, 1999.)

Es gilt

$$\lim_{t \rightarrow 0^+} \frac{f(x^* + td) - f(x^*)}{t} = \nabla f(x^*)^\top d < 0$$

für die Richtungsableitung von f in x entlang d . Daher ist für hinreichend kleine Werte $t > 0$

$$\frac{f(x^* + td) - f(x^*)}{t} < 0.$$

□

Algorithmus A.1 *Ein allgemeines Abstiegsverfahren zur Bestimmung lokaler Minima*

Input:

- Eine Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$.
- Ein Punkt $x^0 \in \mathbb{R}^n$.
- ε mit $\varepsilon > 0$.

Output:

- Ein Punkt $\hat{x} \in \mathbb{R}^n$ mit $\nabla f(\hat{x}) = 0$.
-

- 1: Setze $k := 0$.
- 2: **if** $\|\nabla f(x^k)\|_2 \leq \varepsilon$ **then**
- 3: **return** x^k
- 4: **end if**

Anhang A. Appendix

- 5: Berechne eine Abstiegsrichtung d^k von f in x^k .
- 6: Berechne eine Schrittweite $t^k > 0$, sodass $f(x^k + t^k d^k) < f(x^k)$ gilt.
- 7: Setze $x^{k+1} := x^k + t^k d^k$.
- 8: Setze $k := k + 1$.
- 9: **goto** Zeile 2.

Dieses Verfahren erweckt zwar den Eindruck, nur darauf abzuzielen, einen stationären Punkt von f aufzuspüren, doch es ist zu beachten, dass durch die Verwendung von Abstiegsrichtungen verhindert wird, in einem (lokalen) Maximum zu landen. Es ist allerdings nicht ausgeschlossen, dass die resultierende Lösung ein Sattelpunkt von f ist. Eine Konvergenzanalyse zu Algorithmus A.1 unter bestimmten Voraussetzungen an die Abstiegsrichtungen und die Schrittweiten ist in Geiger und Kanzow (1999) zu finden.

Wir betrachten nun noch folgende Schrittweitenregelung, welche erstmals von Armijo (1966) vorgestellt wurde und daher auch **Armijo-Regel** genannt wird:

Es seien für Algorithmus A.1 zwei feste Werte $\delta \in (0, 1)$ und $\tau \in (0, 1)$ gegeben. Dann wird die Schrittweite t^k in der $(k+1)$ -ten Iteration des Algorithmus als $t^k := \max\{\tau^l \mid l = 0, 1, \dots\}$ mit

$$f(x^k + t^k d^k) \leq f(x^k) + \delta t^k \nabla f(x^k)^\top d^k$$

bestimmt. Dass diese Schrittweitenbestimmung wohldefiniert ist, zeigt folgender Satz:

Satz A.5.2. *Es seien $f : \mathbb{R}^n \rightarrow \mathbb{R}$ stetig differenzierbar, $\delta \in (0, 1)$, $\tau \in (0, 1)$, $x \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ und $\nabla f(x)^\top d < 0$. Dann existiert ein endlicher Wert $l \in \{0, 1, \dots\}$, sodass*

$$f(x + \tau^l d) \leq f(x) + \delta \tau^l \nabla f(x)^\top d.$$

Beweis. (Nach Geiger und Kanzow, 1999.)

Falls für alle $l \in \{0, 1, \dots\}$

$$f(x + \tau^l d) > f(x) + \delta \tau^l \nabla f(x)^\top d \quad \Rightarrow \quad \frac{f(x + \tau^l d) - f(x)}{\tau^l} > \delta \nabla f(x)^\top d$$

gilt, so folgt mit $l \rightarrow \infty$ aus der Definition der Richtungsableitung sowie $\nabla f(x)^\top d < 0$

$$\nabla f(x)^\top d \geq \delta \nabla f(x)^\top d \quad \Rightarrow \quad \delta \geq 1.$$

Dies ist jedoch ein Widerspruch zu $\delta \in (0, 1)$. □

A.6. Newton-Methode

(Siehe Cornuéjols und Tütüncü, 2007.)

Wir betrachten eine stetig differenzierbare Funktion $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ mit

$$F(x) = (f_1(x), \dots, f_n(x))^\top$$

wobei $f_i : \mathbb{R}^n \rightarrow \mathbb{R}, \forall i = 1, \dots, n$. Es ist nun unser Ziel, einen Punkt $x^* \in \mathbb{R}^n$ zu finden, sodass

$$F(x^*) = 0$$

erfüllt ist. Die (mehrdimensionale) Newton-Methode versucht, diese Gleichung folgendermaßen iterativ zu lösen:

Das Verfahren startet in einem vorgegebenen Punkt x^0 und führt in der $(k+1)$ -ten Iteration eine Taylorapproximation 1. Ordnung der Funktion F um den aktuellen Punkt x^k durch:

$$F(x^k + \Delta x^k) \approx F(x^k) + \mathcal{J}_F(x^k) \Delta x^k$$

$\mathcal{J}_F(x^k)$ bezeichnet dabei die Jacobi-Matrix der Funktion F im Punkt x^k . Anschließend wird eine Nullstelle der linearen Approximation von F ermittelt. Es ist also notwendig, das lineare Gleichungssystem

$$\mathcal{J}_F(x^k) \Delta x^k = -F(x^k)$$

zu lösen. Für den Fall, dass $\mathcal{J}_F(x^k)$ regulär ist, resultiert klarerweise die eindeutige Lösung

$$\Delta x^k = -\mathcal{J}_F(x^k)^{-1} F(x^k) \tag{A.1}$$

und wir erhalten den neuen Iterationspunkt

$$x^{k+1} := x^k + \Delta x^k = x^k - \mathcal{J}_F(x^k)^{-1} F(x^k).$$

Damit lässt sich das Newton-Verfahren folgendermaßen beschreiben:

Algorithmus A.2 *Das (lokale) Newton-Verfahren zur Bestimmung von Nullstellen*

Input:

- Eine Funktion $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$.
- Ein Punkt $x^0 \in \mathbb{R}^n$.
- ε mit $\varepsilon > 0$.

Output:

- Eine approximative Lösung $\hat{x} \in \mathbb{R}^n$ des (nichtlinearen) Gleichungssystems $F(x) = 0$.
-

- 1: Setze $k := 0$.
- 2: **if** $\|F(x^k)\|_2 \leq \varepsilon$ **then**

Anhang A. Appendix

- 3: **return** x^k
- 4: **end if**
- 5: Berechne Δx^k als Lösung des (linearen) Gleichungssystems $\mathcal{J}_F(x^k)\Delta x^k = -F(x^k)$.
- 6: Setze $x^{k+1} := x^k + \Delta x^k$.
- 7: Setze $k := k + 1$.
- 8: **goto** Zeile 2.

Geiger und Kanzow (2002) formulieren nachstehenden Satz, welcher Aussagen über die lokale Konvergenz des Newton-Verfahrens trifft:

Satz A.6.1. *Es sei eine stetig differenzierbare Funktion $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ gegeben. $x^* \in \mathbb{R}^n$ erfülle $F(x^*) = 0$ und die Jacobi-Matrix $\mathcal{J}_F(x^*)$ sei regulär. Dann existiert ein Wert $\varepsilon > 0$, sodass für jeden Startpunkt $x^0 \in B_{x^*}(\varepsilon) := \{x \mid \|x - x^*\|_2 < \varepsilon\}$ folgende Aussagen gelten:*

- i) $\mathcal{J}_F(x)$ ist für alle $x \in B_{x^*}(\varepsilon)$ regulär.
- ii) Die von Algorithmus A.2 erzeugte Folge $\{x^k\}_{k=0,1,\dots}$ ist wohldefiniert und konvergiert gegen x^* mit superlinearer Konvergenzrate.²⁹
- iii) Falls $\mathcal{J}_F(x)$ lokal Lipschitz-stetig ist, so ist die Konvergenzrate von Algorithmus A.2 quadratisch.³⁰

Beweis. Siehe Geiger und Kanzow (2002). □

A.7. Quasi-Newton-Methode

(Siehe Geiger und Kanzow (1999), Sun und Yuan (2006) sowie Nocedal und Wright (2006).)

Wir sind wieder (wie in Abschnitt A.5) daran interessiert, ein Minimum x^* einer (nun *zweimal* stetig differenzierbaren) Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$; $x \mapsto f(x)$ zu berechnen. Nehmen wir an, dass $\nabla^2 f(x)$ auf ganz \mathbb{R}^n positiv definit (und damit auch invertierbar) ist, so ist f nach Satz 1.2.2 strikt konvex und die notwendige Bedingung $\nabla f(x^*) = 0$ an den Gradienten von f ist hinreichend. Wir können also das Newton-Verfahren des vorhergehenden Abschnittes mit $F(x) := \nabla f(x)$ anwenden. Gleichung (A.1) nimmt die Form

$$\Delta x^k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k) \tag{A.2}$$

²⁹**Superlineare Konvergenzrate:** Es existiert eine Nullfolge $\{\varepsilon_k\}_{k=0,1,\dots}$ ($\varepsilon_k > 0, \forall k \in \mathbb{N}_0$), sodass

$$\|x^{k+1} - x^*\|_2 \leq \varepsilon_k \|x^k - x^*\|_2, \forall k \in \{0, 1, \dots\}.$$

³⁰**Quadratische Konvergenzrate:** Es existiert eine Konstante $c > 0$ mit

$$\|x^{k+1} - x^*\|_2 \leq c \|x^k - x^*\|_2^2, \forall k \in \{0, 1, \dots\}.$$

Anhang A. Appendix

an, wobei auch $\nabla^2 f(x)^{-1}$ auf ganz \mathbb{R}^n positiv definit ist. Solange $\nabla f(x^k) \neq 0$ gilt und damit noch keine Lösung vorliegt, ist

$$\nabla f(x^k)^\top \Delta x^k = -\nabla f(x^k)^\top \nabla^2 f(x^k)^{-1} \nabla f(x^k) < 0$$

und Δx^k wegen Lemma A.5.1 eine Abstiegsrichtung von f in x^k . Würden wir in Algorithmus A.2 (statt der dort implizit verwendeten Schrittweite $t^k = 1$) noch eine Schrittweitenregelung (wie beispielsweise die Armijo-Regel) einbauen, so entspräche der Newton-Algorithmus einem allgemeinen Abstiegsverfahren (vgl. Algorithmus A.1).

Wenden wir diese Vorgangsweise hingegen für nicht-konvexe Funktionen f an, so ist der in (A.2) definierte Vektor Δx^k unter der Annahme, dass $\nabla^2 f(x^k)$ überhaupt regulär ist, im Allgemeinen keine Abstiegsrichtung mehr. Quasi-Newton-Methoden haben nun zum Ziel, $\nabla^2 f(x^k)$ in jeder Iteration „passend“ durch eine symmetrische Matrix $L_k \succ 0$ zu approximieren, womit $\Delta x^k = -L_k^{-1} \nabla f(x^k)$ wieder eine Abstiegsrichtung ist und im Falle einer Konvergenz gegen x^* mit $\nabla f(x^*) = 0$ zumindest verhindert werden kann, dass x^* ein (lokales) Maximum von f ist.

Eine Möglichkeit, diese Approximationen der Hesse-Matrizen zu bestimmen, besteht laut Nocedal und Wright (2006) darin, nach Bestimmung des Iterationspunktes x^{k+1} eine Taylorapproximation 1. Ordnung von $\nabla f(x)$ um x^{k+1} durchzuführen:

$$\nabla f(x) \approx \nabla f(x^{k+1}) + \nabla^2 f(x^{k+1})(x - x^{k+1})$$

Wir ersetzen nun $\nabla^2 f(x^{k+1})$ durch die approximative Matrix L_{k+1} und fordern, dass

$$\nabla f(x) = \nabla f(x^{k+1}) + L_{k+1}(x - x^{k+1}) \tag{A.3}$$

für die letzten beiden Iterationspunkte x^k und x^{k+1} erfüllt ist. Da (A.3) in $x = x^{k+1}$ ohnehin gilt, reduziert sich dies mit $v^k := x^{k+1} - x^k$ und $w^k := \nabla f(x^{k+1}) - \nabla f(x^k)$ auf die folgende Vorschrift, welche wir *Quasi-Newton-Gleichung* nennen (vgl. Geiger und Kanzow, 1999):

$$L_{k+1}v^k = w^k \tag{A.4}$$

Durch Multiplikation von (A.4) mit $(v^k)^\top$ von links ist unmittelbar ersichtlich, dass eine Matrix L_{k+1} , die (A.4) genügt, nur dann positiv definit sein kann, wenn $(v^k)^\top w^k > 0$ ist. Diese Bedingung ist für die Existenz einer solchen Matrix jedoch nicht nur notwendig, sondern sogar hinreichend:

Lemma A.7.1. *Für zwei Vektoren $v^k \in \mathbb{R}^n \setminus \{0\}$ und $w^k \in \mathbb{R}^n$ existiert genau dann eine symmetrische Matrix $L_{k+1} \in \mathbb{R}^{n \times n}$ mit $L_{k+1} \succ 0$ und $L_{k+1}v^k = w^k$, wenn $(v^k)^\top w^k > 0$ gilt.*

Beweis. Siehe Geiger und Kanzow (1999). □

Bevor wir eine konkrete Approximation der Hesse-Matrix angeben können, benötigen wir noch folgende Aussage:

Anhang A. Appendix

Lemma A.7.2. Für jede symmetrische Matrix $L \in \mathbb{R}^{n \times n}$ mit $L \succcurlyeq 0$ existiert eine eindeutige Matrix $L^{1/2} \in \mathbb{R}^{n \times n}$, die ebenfalls symmetrisch und positiv semidefinit ist und darüber hinaus der Bedingung $L^{1/2}L^{1/2} = L$ genügt. Für $L \succ 0$ ist auch $L^{1/2} \succ 0$. $L^{1/2}$ heißt Quadratwurzel von L .

Beweis. (Vgl. Kanzow, 2005.)

Die reelle, symmetrische und positiv semidefinite Matrix L ist normal (d.h. $L^\top L = LL^\top$) und deswegen nach Kanzow (2005) unitär diagonalisierbar mit $U^\top L U = D$. $U \in \mathbb{R}^{n \times n}$ ist regulär und orthogonal und $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ eine Diagonalmatrix mit den Eigenwerten $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ von L auf der Diagonalen. Definieren wir nun $D^{1/2} := \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$ und $L^{1/2} := U D^{1/2} U^\top$, so ist $L^{1/2}$ offenbar symmetrisch, positiv semidefinit und $L^{1/2}L^{1/2} = L$. Für eine positiv definite Matrix L sind alle Eigenwerte strikt positiv, weshalb auch die Quadratwurzel $L^{1/2}$ positiv definit ist. Die Eindeutigkeit von $L^{1/2}$ ist in Kanzow (2005) nachzulesen. \square

Ziel der Quasi-Newton-Verfahren ist nun, mit einer positiv definiten und symmetrischen Matrix L_0 zu starten und diese in jedem Schritt iterativ „fortzuschreiben“. Zwei populäre Methoden, welche auf (A.4) basieren, sind die PSB-Formel (**P**owell-**S**ymmetric-**B**royden-Formel) sowie die DFP-Formel (**D**avidon-**F**letcher-**P**owell-Formel). Eine andere Variante, die nach **B**royden (1970), **F**letcher (1970), **G**oldfarb (1970) und **S**hanno (1970) benannte BFGS-Formel, leitet sich hingegen von der *inversen Quasi-Newton-Gleichung* ab:

$$L_{k+1}^{-1} w^k = v^k \tag{A.5}$$

Der nachstehende Satz, welcher aus Geiger und Kanzow (1999) stammt, gibt nun die Aufdatierung einer vorliegende Matrix L_k^{-1} mithilfe der BFGS-Formel an, wobei die dort auftretende *Frobenius-Norm* für eine Matrix $M = (M_{i,j})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} \in \mathbb{R}^{m \times n}$ wie folgt definiert ist:

$$\|M\|_F := \sqrt{\sum_{i=1}^m \sum_{j=1}^n M_{i,j}^2}$$

Satz A.7.3. (*BFGS-Formel*)

Es seien zwei Vektoren $v^k, w^k \in \mathbb{R}^n$ mit $(v^k)^\top w^k > 0$ sowie eine symmetrische Matrix $L_k \in \mathbb{R}^{n \times n}$ mit $L_k \succ 0$ gegeben. $W_{k+1} \in \mathbb{R}^{n \times n}$ sei eine beliebige symmetrische Matrix mit $W_{k+1} \succ 0$ sowie $W_{k+1} v^k = w^k$ und $W_{k+1}^{1/2}$ sei die Quadratwurzel von W_{k+1} . Dann gelten folgende Aussagen:

a) Das Optimierungsproblem

$$\begin{aligned} \min_{L^{-1} \in \mathbb{R}^{n \times n}} \quad & \left\| W_{k+1}^{1/2} (L^{-1} - L_k^{-1}) W_{k+1}^{1/2} \right\|_F^2 \\ \text{s.t.} \quad & L^{-1} w^k = v^k \\ & (L^{-1})^\top = L^{-1} \end{aligned}$$

Anhang A. Appendix

besitzt die eindeutig bestimmte Optimallösung

$$L_{k+1}^{-1} := L_k^{-1} + \frac{(v^k - L_k^{-1}w^k)(v^k)^\top + v^k(v^k - L_k^{-1}w^k)^\top}{(w^k)^\top v^k} - \frac{(v^k - L_k^{-1}w^k)^\top w^k}{((w^k)^\top v^k)^2} v^k (v^k)^\top.$$

b) Die Inverse der Matrix L_{k+1}^{-1} aus a) ist

$$L_{k+1} := L_k + \frac{w^k(w^k)^\top}{(v^k)^\top w^k} - \frac{L_k v^k (v^k)^\top L_k}{(v^k)^\top L_k v^k}. \quad (\text{A.6})$$

Beweis. Siehe Geiger und Kanzow (1999) für Teil a). Dabei liefern die Lemmata A.7.1 und A.7.2 die Existenz von W_{k+1} bzw. $W_{k+1}^{1/2}$ entsprechend den Anforderungen des Satzes. Der Beweis von Teil b) ist mithilfe der Sherman-Morrison-Woodbury-Formel möglich (vgl. Sun und Yuan, 2006). \square

Abbildungsverzeichnis

2.1. Die effiziente Front des „Mean-Variance“-Modells.	20
5.1. Farbcodes der 63 Aktien aus dem DJCA	109
5.2. DJCA: „Mean-Variance“-Portfolios des Zeitraumes 2004/2005	112
5.3. DJCA: approximative „Risk-Parity“-Portfolios des Zeitraumes 2004/2005	112
5.4. DJIA: „Mean-Variance“-Portfolios des Zeitraumes 2004/2005	113
5.5. DJIA: approximative „Risk-Parity“-Portfolios des Zeitraumes 2004/2005	113
5.6. DJTA: „Mean-Variance“-Portfolios des Zeitraumes 2004/2005	114
5.7. DJTA: approximative „Risk-Parity“-Portfolios des Zeitraumes 2004/2005	114
5.8. DJUA: „Mean-Variance“-Portfolios des Zeitraumes 2004/2005	115
5.9. DJUA: approximative „Risk-Parity“-Portfolios des Zeitraumes 2004/2005	115
5.10. DJCA: „Mean-Variance“- und approx. „Risk-Parity“-Portfolios von 2004/2005 im Vergleich	116
5.11. DJCA: „Mean-Variance“-Portfolios des Zeitraumes 2008/2009	118
5.12. DJCA: approximative „Risk-Parity“-Portfolios des Zeitraumes 2008/2009	118
5.13. DJIA: „Mean-Variance“-Portfolios des Zeitraumes 2008/2009	119
5.14. DJIA: approximative „Risk-Parity“-Portfolios des Zeitraumes 2008/2009	119
5.15. DJTA: „Mean-Variance“-Portfolios des Zeitraumes 2008/2009	120
5.16. DJTA: approximative „Risk-Parity“-Portfolios des Zeitraumes 2008/2009	120
5.17. DJUA: „Mean-Variance“-Portfolios des Zeitraumes 2008/2009	121
5.18. DJUA: approximative „Risk-Parity“-Portfolios des Zeitraumes 2008/2009	121
5.19. DJCA: „Mean-Variance“- und approx. „Risk-Parity“-Portfolios von 2008/2009 im Vergleich	122

Tabellenverzeichnis

5.1. Aktien des DJCA sowie seiner Teilindizes DJIA, DJTA und DJUA	100
5.2. Berechnung eines „Risk-Parity“-Portfolios im „Long-Only“-Fall für den DJCA	103
5.3. Berechnung eines „Risk-Parity“-Portfolios mit „Long“ und „Short Positions“ für den DJCA	104
5.4. Berechnung eines „Minimum-Variance“-Portfolios für den DJCA	105
5.5. Varianzen der Aktien des DJCA und seiner Teilindizes für die einzelnen Jahre von 2003 bis 2014	107

Literatur

- Alt, Walter (2010). *Nichtlineare Optimierung*. Vieweg+Teubner Verlag.
- Armijo, Larry (1966). „Minimization of functions having Lipschitz continuous first partial derivatives“. In: *Pacific Journal of Mathematics* 16 (1), S. 1–3.
- Bai, Xi, Katya Scheinberg und Reha Tütüncü (2013). *Least-squares approach to risk parity in portfolio selection*. URL: http://www.optimization-online.org/DB_HTML/2013/10/4089.html (besucht am 23.02.2015).
- Best, Michael J. und Robert R. Grauer (1991). „On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results“. In: *The Review of Financial Studies* 4 (2), S. 315–342.
- Brito, R. P. und L. N. Vicente (2012). *Efficient Cardinality/Mean-Variance Portfolios*. URL: http://www.optimization-online.org/DB_HTML/2012/03/3381.html (besucht am 23.02.2015).
- Broyden, Charles George (1970). „The convergence of a class of double-rank minimization algorithms“. In: *Journal of the Institute of Mathematics and Its Applications* 6, S. 76–90.
- Bruder, Benjamin und Thierry Roncalli (2012). *Managing Risk Exposures using the Risk Budgeting Approach*. URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2009778 (besucht am 23.02.2015).
- Cornuéjols, Gérard und Reha Tütüncü (2007). *Optimization Methods in Finance*. Cambridge University Press.
- Curtis, Frank und Jorge Nocedal (2007). „Steeplength selection in interior-point methods for quadratic programming“. In: *Applied Mathematics Letters* 20 (5), S. 516–523.
- D’Apuzzo, Marco, Valentina De Simone und Daniela di Serafino (2010). „On mutual impact of numerical linear algebra and large-scale optimization with focus on interior point methods“. In: *Computational Optimization and Applications* 45 (2), S. 283–310.
- Fletcher, Roger (1970). „A New Approach to Variable Metric Algorithms“. In: *Computer Journal* 13 (3), S. 317–322.
- Geiger, Carl und Christian Kanzow (1999). *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*. Springer.
- Geiger, Carl und Christian Kanzow (2002). *Theorie und Numerik restringierter Optimierungsaufgaben*. Springer.
- Goldfarb, Donald (1970). „A Family of Variable Metric Updates Derived by Variational Means“. In: *Mathematics of Computation* 24 (109), S. 23–26.
- Gondzio, Jacek (2013). „Convergence Analysis of an Inexact Feasible Interior Point Method for Convex Quadratic Programming“. In: *SIAM Journal on Optimization* 23 (3), S. 1510–1527.

Literatur

- Jarre, Florian und Josef Stoer (2004). *Optimierung*. Springer.
- Kanzow, Christian (2005). *Numerik linearer Gleichungssysteme: Direkte und iterative Verfahren*. Springer.
- Kaya, Hakan und Wai Lee (2012). „Demystifying Risk Parity“. In: *Neuberger Berman*.
- Luenberger, David und Yinyue Ye (2008). *Linear and Nonlinear Programming*. Springer.
- Maillard, Sébastien, Thierry Roncalli und Jérôme Teiletche (2010). „On the properties of equally-weighted risk contributions portfolios“. In: *The Journal of Portfolio Management* 36 (4), S. 60–70.
- Maratos, Nicholas (1978). *Exact penalty function algorithms for finite dimensional and control optimization problems*. Ph.D. Thesis.
- Markowitz, Harry M. (1952). „Portfolio Selection“. In: *The Review of Financial Studies* 7 (1), S. 77–91.
- Markowitz, Harry M. (1987). *Mean-Variance Analysis in Portfolio Choice and Capital Markets*. Basil Blackwell.
- Mehrotra, Sanjay (1992). „On the Implementation of a Primal-Dual Interior Point Method“. In: *SIAM Journal on Optimization* 2 (4), S. 575–601.
- Merton, Robert C. (1980). „On Estimating the Expected Return on the Market: An Exploratory Investigation“. In: *Journal of Financial Economics* 8, S. 323–361.
- Nocedal, Jorge und Stephen J. Wright (2006). *Numerical Optimization*. Springer.
- Powell, Michael James David (1978). „A fast algorithm for nonlinearly constrained optimization calculations“. In: *Lecture Notes in Mathematics* 630, S. 144–157.
- Qian, Edward (2005). *Risk Parity Portfolios, Research Paper, PanAgora*. URL: <https://www.panagora.com/assets/PanAgora-Risk-Parity-Portfolios-Efficient-Portfolios-Through-True-Diversification.pdf> (besucht am 23.02.2015).
- Roncalli, Thierry (2013). *Introduction to Risk Parity and Budgeting*. Chapman & Hall/CRC Financial Mathematics Series. Taylor & Francis.
- Schade, Philipp (2008). *Innere-Punkte-Verfahren mit Redundanzerkennung für die Quadratische Optimierung*. Springer Gabler.
- Shanno, David F. (1970). „Conditioning of quasi-Newton methods for function minimization“. In: *Mathematics of Computation* 24 (111), S. 647–656.
- Spellucci, Peter (1993). *Numerische Verfahren der nichtlinearen Optimierung*. Birkhäuser Verlag.
- Spinu, Florin (2013). *An Algorithm for Computing Risk Parity Weights*. URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2297383 (besucht am 23.02.2015).
- Steinbach, Marc C. (2001). „Markowitz Revisited: Mean-Variance Models in Financial Portfolio Analysis“. In: *SIAM Review* 43 (1), S. 31–85.
- Sun, Wenyu und Ya-Xiang Yuan (2006). *Optimization Theory and Methods: Nonlinear Programming*. Springer.
- Werner, Dirk (2011). *Funktionalanalysis*. Springer.
- Wright, Stephen J. (1997). *Primal-Dual Interior-Point Methods*. SIAM.