Universität Innsbruck



Institut für Mathematik Angewandte Mathematik

Das stochastische Gradientenverfahren

Chris Wendler Chris.Wendler@student.uibk.ac.at

Bachelorarbeit

Betreuer: Markus Haltmeier Markus.Haltmeier@uibk.ac.at 22. Mai 2017

Kurzzusammenfassung

Gradientenverfahren sind iterative Optimierungsverfahren, die aufgrund ihrer einfachen Struktur und Flexibilität auf ein breites Spektrum von Zielfunktionen angewendet werden können. Unglücklicherweise ist das traditionelle Gradientenverfahren langsam, wenn die Berechnung des Gradienten der Zielfunktion aufwändig ist. Deshalb ist es für eine große Familie von praxisrelevanten Zielfunktionen, zum Beispiel für Zielfunktionen die sich als Linearkombination von einer großen Anzahl von Teilfunktionen zusammensetzen, nicht optimal. Bei Zielfunktionen vom genannten Typ kann das Gradientenverfahren durch stochastische Approximationen des Gradienten beschleunigt werden. Wird der Gradient durch den Gradienten einer zufällig gewählten Teilfunktion ersetzt, so spricht man vom stochastischem Gradientenverfahren, welches sublinear konvergiert. In dieser Arbeit werden wir eine linear konvergierende Variante des stochastischen Gradientenverfahrens betrachten. Im Gegensatz zur klassischen stochastischen Gradientenmethode werden bei dieser Variante die vorigen Gradienten der Teilfunktionen beibehalten. Diese Variante des stochastischen Gradientenverfahrens findet unter anderem im Rahmen des maschinellen Lernens viele Anwendungen. In dieser Arbeit werden wir anhand von numerischen Experimenten zur zirkulären Radontransformation zeigen, dass das Verfahren auch zum Lösen von schlecht konditionierten Gleichungssystemen gut geeignet ist und bessere Ergebnisse als konventionelle Verfahren liefert.

Inhaltsverzeichnis

Kι	urzzusammenfassung	i				
In	haltsverzeichnis	iii				
Al	bbildungsverzeichnis	\mathbf{v}				
Ta	abellenverzeichnis	vii				
Er	rklärung	ix				
1	Einleitung	1				
2	Hintergrund	3				
	2.1 Konvergenzbegriffe	3				
	2.2 Vereinfachende Annahmen	4				
	2.3 Gradientenverfahren	5				
	2.3.1 Konvergenzanalyse	5				
	2.3.2 Schrittkosten	6 6				
	2.4 Stochastisches Grädientenverfahren	07				
	2.4.1 Schnittkösten	$\frac{1}{7}$				
3	Das Stochastic-Average-Gradient Verfahren	11				
	3.1 Schrittkosten	11				
	3.2 Theoretische Konvergenzanalyse	12				
	3.2.1 Problemformulierung und Notation	12				
	3.2.2 Beweis nach Roux et al. (2012)	13				
4	Numerische Ergebnisse					
	4.1 Zirkuläre Radontransformation	21				
	4.1.1 Aufbau und Problemformulierung	22				
	4.1.2 Landweber-Kaczmarz Verfahren	22				
	4.1.3 Numerische Experimente	23				
5	Konklusion	29				
\mathbf{Li}	teraturverzeichnis	31				

Abbildungsverzeichnis

4.1 Schematische Darstellung der zirkulären Radontransformation. Die Sensore			
	entlang eines Kreises um das Objekt f angeordnet. Für jede Position $a = (x_0, y_0)$		
	auf dem Kreis und jeden Radius r liefert die zirkuläre Radontransformation Rf		
	einen Wert $Rf(a,r)$	22	
4.2	Vergleich der vorgestellten Methoden bezüglich der Fehler der Iterierten und		
	der Fehler der Zielfunktion am Beispiel der zirkulären Radontransformation. Die		
	y-Achse ist logarithmisch skaliert. Die x -Achse ist in Zyklen unterteilt. Ein Zy-		
	klus entspricht dabei einem Durchlauf durch die Daten, also einer Iteration des		
	Gradientenverfahrens oder n Iterationen der anderen Verfahren	27	
4.3	Vergleich der vorgestellten Methoden bis eine gewisse Genauigkeit erreicht ist. Die		
	Abbildung stellt die Anzahl der Zyklen dar, die erforderlich sind um eine gewisse		
	Genauigkeit zu erreichen.	28	

Tabellenverzeichnis

4.1	Vergleich der Iterationskosten der verschiedenen Verfahren in Millisekunden. Die	
	Tabelle enthält die über 50 Zyklen gemittelten Iterationszeiten der verschiedenen	
	Verfahren	24
4.2	Illustration der Startphase der approximativen Verfahren. Zwischen den einzel-	
	nen Bildern liegen jeweils fünf Iterationen der zugehörigen Verfahren. Für jedes	
	Verfahren werden also die Rekonstruktionen von Iteration 1 bis 196 dargestellt. $$.	26

Erklärung

Ich erkläre hiermit an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Alle Stellen, die wörtlich oder inhaltlich den angegebenen Quellen entnommen wurden, sind als solche kenntlich gemacht.

Signiert: Datum:

Kapitel 1

Einleitung

In dieser Arbeit werden Optimierungsprobleme der Form

betrachtet. Die Zielfunktion

$$x \mapsto g(x) = \frac{1}{n} \sum_{i=1}^{n} f_i(x) \tag{1.2}$$

ist dabei eine endliche Linearkombination von Teilfunktionen und von großer Praxisrelevanz. Zielfunktionen dieser Form treten besonders häufig im statistischen Kontext und beim maschinellen Lernen auf, wo oft versucht wird eine unbekannte Funktion anhand einer endlichen Menge von Ein-/Ausgabepaaren zu rekonstruieren. Dabei wäre die *i*-te Teilfunktion zum Beispiel der Fehler der betrachteten Hypothesen-Funktion ausgewertet am *i*-ten Ein-/Ausgabepaar. Die Funktionsauswertungen $f_i(x)$ können also als Teilfehler und g(x) als Gesamtfehler einer von x parametrisierten Hypothese interpretiert werden. Des Weiteren tritt Optimierungsproblem (1.1) beim Lösen von Gleichungssystemen auf, wo eine ähnliche Interpretation möglich ist. Dazu betrachten wir ein Beispiel:

Beispiel 1. (Gleichungssysteme lösen) Sei $A \in \mathbb{R}^{n \times p}$, $b \in \mathbb{R}^n$ und sei $n \ge p$. Aufgrund der positiven Definitheit der Norm kann die Lösung des überbestimmten Gleichungssystems Ax = b als

$$x^* = \underset{x \in \mathbb{R}^p}{\arg\min} \frac{1}{2n} \|Ax - b\|^2$$
(1.3)

geschrieben werden. Entfaltet man hierbei die Definition der Norm, so erhält man eine Zielfunktion der Form (1.2)

$$g(x) = \frac{1}{2n} ||Ax - b||^2$$

= $\frac{1}{n} \sum_{i=1}^n \frac{1}{2} |A_{i-x} - b_i|^2,$ (1.4)

wobei $A_{i-} \in \mathbb{R}^{1 \times p}$ die i-te Zeile der Matrix A bezeichnet. Beim Lösen von Gleichungssystemen kann die Funktionsauswertung $f_i(x)$ also auch als Teilfehler des Lösungskandidaten x interpretiert werden.

Um Optimierungsprobleme der Form (1.1) zu lösen, werden in dieser Arbeit verschiedene Gradientenverfahren betrachtet. Dabei werden wir besonderes Augenmerk auf stochastische Varianten des Gradientenverfahrens legen. Gradientenverfahren sind iterative Optimierungsverfahren, welche aufgrund ihrer Einfachheit in der Praxis sehr beliebt sind. Im Gegensatz zu komplexeren Verfahren, zum Beispiel dem Newton-Verfahren, wird zur Anwendung von Gradientenverfahren lediglich die einfache Differenzierbarkeit der Zielfunktion benötigt.

Der Hauptteil dieser Arbeit besteht aus der numerischen und theoretischen Konvergenzanalyse einiger ausgewählter Verfahren, nämlich dem traditionellen Gradientenverfahren, dem stochastischen Gradientenverfahren und dem von (Roux et al., 2012) eingeführten Stochastic-Average-Gradient Verfahren (SAG-Verfahren). Bemerkenswert ist dabei der Vergleich der Verfahren hinsichtlich ihrer Iterationskosten und Konvergenzgeschwindigkeit. Dabei stellt sich heraus, dass das traditionelle Gradientenverfahren hohe Schrittkosten hat, dafür sehr schnell konvergiert. Um den hohen Schrittkosten entgegenzuwirken wird das Gradientenverfahren beim stochastischen Gradientenverfahren approximiert. Die Approximation führt zwar zu geringen Schrittkosten, allerdings geschieht dies auf Kosten der Konvergenzgeschwindigkeit. Auf natürliche Weise stellt sich die Frage ob die Vorzüge beider Verfahren vereint werden können. Diese wird vom SAG-Verfahren, welches schnell konvergiert und billige Schritte hat, bejaht.

Kapitel 2

Hintergrund

2.1 Konvergenzbegriffe

Definition 2. (Iteratives Optimierungsverfahren) Ein iteratives Optimierungsverfahren zur Lösung eines Optimierungsproblems

$$\min_{x \in \mathbb{R}^p} g(x) \tag{2.1}$$

ist durch eine Schrittfunktion

$$\Phi: \mathbb{R}^p \to \mathbb{R}^p: x \mapsto \Phi(x), \tag{2.2}$$

welche die Iterierte x_k bezüglich des Optimierungsproblems verbessern soll, gegeben. Die Iterierten sind durch die Rekursion

$$x^{k+1} = \Phi(x^k),$$

$$x^0 \in \mathbb{R}^p, k \in \mathbb{N}$$
(2.3)

definiert, wobei x^0 den Startwert bezeichnet. Konvergiert das Verfahren gegen eine Minimalstelle von g, so kann iterativ die Lösung eines Optimierungsproblems gefunden werden.

Zur Beurteilung iterativer Optimierungsverfahren wird unter anderem die Konvergenzgeschwindigkeit, mit der die Folge der Iterierten $(x^k)_{k \in \mathbb{N}}$, gegen eine Lösung x^* des Optimierungsproblems konvergiert, verwendet. Dabei unterscheidet man zwischen folgenden Typen der Konvergenz.

Definition 3. (Konvergenzgeschwindigkeit nach Reinhardt et al. (2013)) Es seien $(x^k)_{k \in \mathbb{N}} \in (\mathbb{R}^p)^{\mathbb{N}}$ und $\lim_{k \to \infty} x^k = x^*$. Die Folge $(x^k)_{k \in \mathbb{N}}$ konvergiert gegen x^*

• Q-sublinear, wenn eine Folge $(c_k)_{k \in \mathbb{N}}$ mit $\lim_{k \to \infty} c_k = 1$ und ein $k_0 \in \mathbb{N}$ existieren, sodass

$$\|x^{k+1} - x^*\| \le c_k \|x^k - x^*\|$$

• Q-linear mit dem Konvergenzfaktor $C \in (0,1)$, wenn ein $k_0 \in \mathbb{N}$ existiert, sodass

$$\|x^{k+1} - x^*\| \le C \|x^k - x^*\|$$

• Q-superlinear, wenn eine positive Nullfolge $(c_k)_{k \in \mathbb{N}}$ und ein $k_0 \in \mathbb{N}$ existieren, sodass

$$\|x^{k+1} - x^*\| \le c_k \|x^k - x^*\|$$

• Q-quadratisch mit dem Konvergenzfaktor C > 0, wenn ein $k_0 \in \mathbb{N}$ existiert, sodass

$$\|x^{k+1} - x^*\| \le C \|x^k - x^*\|^2$$

für alle $k \in \mathbb{N}$ mit $k \ge k_0$ gilt.

Das Präfix "Q-" verweist dabei auf die Tatsache, dass die obigen Definitionen auf dem Quotientenkriterium zur absoluten Konvergenz von Reihen basieren.

Definition 4. (Q-Konvergenz der Ordnung d) Es seien $p \in \mathbb{N}_{\geq 1}$, $(x^k)_{k \in \mathbb{N}} \in (\mathbb{R}^p)^{\mathbb{N}}$ und $\lim_{k \to \infty} x^k = x^*$. Die Folge $(x^k)_{k \in \mathbb{N}}$ heißt Q-konvergent der Ordnung d, wenn ein C > 0 und ein $k_0 \in \mathbb{N}$ existieren, sodass

$$\|x^{k+1} - x^*\| \le C \|x^k - x^*\|^d \tag{2.4}$$

für alle $k \in \mathbb{N}$ mit $k \ge k_0$ gilt.

Definition 5. (Exponentielle Konvergenz) Es seien $(x^k)_{k \in \mathbb{N}} \in (\mathbb{R}^p)^{\mathbb{N}}$ und $\lim_{k \to \infty} x^k = x^*$. Die Folge $(x^k)_{k \in \mathbb{N}}$ mit dem Grenzwert x^* konvergiert exponentiell, wenn $\alpha \in (0,1)$, $\beta > 0$ und ein $k_0 \in \mathbb{N}$ existieren, sodass

$$\|x^k - x^*\| \le \beta \alpha^k \tag{2.5}$$

für alle $k \in \mathbb{N}$ mit $k \ge k_0$ gilt.

In der Literatur wird für die exponentielle Konvergenz oft diese Kurzschreibweise verwendet:

Bemerkung 6. (Kurzschreibweise) Für eine exponentiell konvergente Folge $(x_k)_{k \in \mathbb{N}}$ gilt

$$\|x^k - x^*\| \in \mathcal{O}(\alpha^k). \tag{2.6}$$

Außerdem werden die Konzepte der Q-linearen Konvergenz und der exponentiellen Konvergenz austauschbar verwendet, was durch das nachfolgende Lemma gerechtfertigt ist.

Lemma 7. Eine Q-linear konvergente Folge $(x^k)_{k \in \mathbb{N}}$ ist auch exponentiell konvergent.

Beweis. Es gelte $||x^{k+1} - x^*|| \le C ||x^k - x^*||$ für $C \in (0, 1)$. Die Behauptung folgt aus der Rekursion

$$\begin{aligned} \|x^{1} - x^{*}\| &\leq C \|x^{0} - x^{*}\| \\ \|x^{2} - x^{*}\| &\leq C \|x^{1} - x^{*}\| \leq C^{2} \|x^{0} - x^{*}\| \\ &\vdots \\ \|x^{k} - x^{*}\| &\leq \underbrace{C^{k}}_{=:\alpha^{k}} \underbrace{\|x^{0} - x^{*}\|}_{=:\beta}. \end{aligned}$$

$$(2.7)$$

-	_	-	-	

2.2 Vereinfachende Annahmen

Im restlichen Teil der Arbeit werden wir immer wieder die Anwendung von verschiedenen Gradientenverfahren auf das Optimierungsproblem

$$\min_{x \in \mathbb{R}^p} \underbrace{\frac{1}{n} \sum_{i=1}^n f_i(x)}_{:=g(x)}$$

studieren. Dabei werden wir der Einfachheit halber annehmen, dass alle auf der offenen Menge $\mathcal{D} \subseteq \mathbb{R}^p$ definierten reellwertigen Teilfunktionen $f_i : \mathcal{D} \to \mathbb{R}$ in Optimierungsproblem (1.1) differenzierbar und dass alle Gradienten ∇f_i Lipschitz-stetig mit Konstante L sind für $i \in \{1, \ldots, n\}$. Des Weiteren nehmen wir an, dass g stark konvex mit der Konstante μ ist.

Definition 8. (Starke Konvexität) Eine reellwertige differenzierbare Funktion $f : \mathcal{D} \subseteq \mathbb{R}^p \to \mathbb{R}$ heißt stark konvex mit der Konstante $\mu > 0$, wenn sie für $x, y \in \mathcal{D}$ die Ungleichung

$$(\nabla f(x) - \nabla f(y))^T (x - y) \ge \mu ||x - y||^2$$
 (2.8)

erfüllt. Eine dazu äquivalente Bedingung ist (siehe Nesterov (2004))

$$f(y) \ge f(x) + \nabla f(x)^T (y - x) + \frac{\mu}{2} ||y - x||^2.$$
(2.9)

2.3 Gradientenverfahren

Wir beginnen mit einem der einfachsten iterativen Verfahren zur Lösung von Optimierungsproblemen, nämlich dem sogenannten Gradientenverfahren. Dem Gradientenverfahren liegt eine einfache Idee zu Grunde: Um der Minimalstelle x^* der Zielfunktion g in jeder Iteration näher zu kommen, wird die Iterierte in die Richtung des größten Abstiegs verschoben. Betrachtet man die Richtungsableitung der Zielfunktion g in Richtung v im Punkt x, welche als Anstieg in Richtung v im Punkt x interpretiert werden kann,

$$\partial_v g(x) = \langle \nabla g(x), v \rangle, \qquad (2.10)$$

so stellt man anhand der Überlegung

$$\langle \nabla g(x), v \rangle = \cos(\gamma) \| \nabla g(x) \| \| v \|, \tag{2.11}$$

in welcher γ den Winkel zwischen $\nabla g(x)$ und v bezeichnet, schnell fest, dass die Richtung des größten Abstiegs durch $-\nabla g(x)$ gegeben ist. Somit erhält man für das Gradientenverfahren folgende Iterationsvorschrift

$$x^{0} \in \mathbb{R}^{p}$$

$$x^{k+1} = x^{k} - \alpha_{k} \nabla g(x^{k})$$

$$= x^{k} - \frac{\alpha_{k}}{n} \sum_{i=1}^{n} \nabla f_{i}(x^{k}),$$
(2.12)

wobei $\alpha_k > 0$ die Schrittweite in der k-ten Iteration bezeichnet. Wir werden sehen, dass das Verfahren für konvexe Zielfunktionen bei geeigneter Wahl der Sequenz der Schrittweiten $(\alpha_k)_{k \in \mathbb{N}}$ mit linearer Geschwindigkeit zur eindeutigen Minimalstelle x^* konvergiert.

2.3.1 Konvergenzanalyse

Wie bereits erwähnt, stellt die Konvergenzgeschwindigkeit eines Verfahrens ein wichtiges Beurteilungskriterium dar. Des Weiteren werden wir die Berechnungskosten für einen einzelnen Schritt des Verfahrens analysieren.

Proposition 9. (Q-lineare Konvergenz des Abstiegsverfahrens nach Nesterov (2004)) Für eine μ -stark konvexe Zielfunktion g mit L-Lipschitz-stetigem Gradienten ∇g und ausreichend kleiner konstanter Schrittweite α konvergiert das Gradientenverfahren mit Schrittweite $\alpha \in (0, \frac{2}{\mu+L})$ Q-linear gegen eine Minimalstelle von g.

Beweis. Die Aussage folgt aus der Überlegung

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &= \|x^k - \alpha \nabla g(x^k) - x^*\|^2 & [\text{Gleichung (2.12)}] \\ &= \|x^k - x^*\|^2 - 2\alpha \left\langle \nabla g(x^k), x^k - x^* \right\rangle + \alpha^2 \|\nabla g(x^k)\|^2 & [\|\cdot\|^2 = \langle \cdot, \cdot \rangle] \\ &\leq \left(1 - \frac{2\alpha\mu L}{\mu + L}\right) \|x^k - x^*\|^2 + \alpha \left(\alpha - \frac{2}{\mu + L}\right) \|\nabla g(x^k)\|^2 & [\text{Theorem 2.1.12} \\ &(\text{Nesterov, 2004)}] \\ &\leq \left(1 - \frac{2\alpha\mu L}{\mu + L}\right) \|x^k - x^*\|^2. \end{aligned}$$

$$(2.13)$$

Dabei bezeichnet L eine Lipschitzkonstante von ∇g und μ , die zur μ -stark konvexen Funktion g gehörige Konstante. Theorem 2.1.12 aus Nesterov (2004) besagt, dass für eine μ -stark konvexe Funktion g mit L-Lipschitz-stetigem Gradienten ∇g und $x, y \in \mathbb{R}^p$ folgende Ungleichung erfüllt ist:

$$\langle \nabla g(x) - \nabla g(y), x - y \rangle \ge \frac{\mu L}{\mu + L} \|x - y\|^2 + \frac{1}{\mu + L} \|\nabla f(x) - \nabla f(y)\|^2.$$
 (2.14)

Für eine konstante Schrittweite $\alpha \in (0, \frac{2}{\mu+L})$ konvergiert das Gradientenverfahren somit Qlinear.

2.3.2 Schrittkosten

Betrachtet nun man die Kosten eines einzelnen Schrittes, welche sich im Wesentlichen aus der Berechnung des Gradienten $\nabla g(x)$ zusammensetzen, so stellt man fest, dass diese für unsere Zielfunktionen $g(x) = \frac{1}{n} \sum_{i=1}^{n} f_i(x)$ von der Anzahl n der Teilfunktionen f_i abhängen.

Betrachtet man zum Beispiel die Zielfunktion, die man beim Lösen von überbestimmten Gleichungssystemen erhält, so gilt

$$g(x) = \frac{1}{2n} \|Ax - b\|^2, A \in \mathbb{R}^{n \times p}$$
$$\nabla g(x) = \frac{1}{n} A^T (Ax - b).$$
(2.15)

und

Die wesentlichen Kosten einer Iteration liegen in $\mathcal{O}(np)$, was den Kosten der Matrix-Vektor Multiplikationen Ax beziehungsweise $A^T(Ax - b)$ entspricht. Die Kosten der Differenzbildung Ax - b sind geringer und liegen in $\mathcal{O}(n)$. In praxisrelevanten Gleichungssystemen können pund n Größenordnungen annehmen, die eine effektive Anwendung des Gradientenverfahrens verhindern.

Auch beim maschinellen Lernen, wo f_i meist den Fehler einer durch x parametrisierten Hypothese am *i*-ten Datenpunkt beschreibt, kann die Anzahl von Teilfunktionen im Millionenbereich liegen. Aufgrund der daraus resultierenden hohen Schrittkosten ist das Gradientenverfahren auch dort oft ungeeignet.

2.4 Stochastisches Gradientenverfahren

Abgesehen von den hohen Schrittkosten bietet das Gradientenverfahren einige attraktive Eigenschaften, wie dessen Einfachheit oder dessen Q-lineare Konvergenzgeschwindigkeit. Daher liegt es nahe die teure Berechnung des Gradienten $\nabla g(x)$ durch eine billigere Operation zu ersetzen, um die Schrittkosten zu reduzieren. Beim stochastischen Gradientenverfahren wird der Gradient $\nabla g(x^k)$ in der Vorschrift (2.12) durch einen Teilgradienten $\nabla f_{i_k}(x^k)$ ersetzt und man erhält die Iterationsvorschrift

$$x^{0} \in \mathbb{R}^{p}$$

$$x^{k+1} = x^{k} - \alpha_{k} \nabla f_{i_{k}}(x^{k}) \quad \text{mit } i_{k} \sim \mathcal{U}_{\{1,\dots,n\}},$$

$$(2.16)$$

wobei der Index i_k einer diskreten Gleichverteilung $\mathcal{U}_{\{1,\dots,n\}}$ auf der Menge $\{1,\dots,n\}$ folgt. Dabei handelt es sich bei $\nabla f_{i_k}(x^k)$ um einen erwartungstreuen Schätzer von $\nabla g(x^k)$, denn es gilt

$$\mathbb{E}[\nabla f_{i_k}(x^k)] = \sum_{i=1}^n \mathbb{P}(i_k = i) \nabla f_i(x^k)$$

$$= \sum_{i=1}^n \mathcal{U}_{\{1,\dots,n\}}(\{i\}) \nabla f_i(x^k)$$

$$= \sum_{i=1}^n \frac{1}{n} \nabla f_i(x^k)$$

$$= \nabla g(x^k).$$

(2.17)

2.4.1 Schrittkosten

Offensichtlich hängt der Iterationsschritt in Vorschrift (2.16) nicht von der Anzahl der Teilfunktionen ab. Des Weiteren benötigt die Generierung einer gleichverteilten Zufallszahl in den meisten Programmiersprachen nur eine konstante Anzahl von Operationen. Daher besteht der Rechenaufwand für eine Iteration aus der Auswertung des Gradienten einer Teilfunktion.

Betrachten wir nun wiederum die Zielfunktion, die beim Lösen von überbestimmten Gleichungssystemen auftritt

$$g(x) = \frac{1}{2n} ||Ax - b||^2$$

= $\frac{1}{2n} \sum_{i=1}^n |A_{i-}^T x - b_i|^2, A \in \mathbb{R}^{n \times p},$

so gilt

$$\nabla g(x) = \frac{1}{n} A^T (Ax - b) \tag{2.18}$$

und

$$\nabla f_i(x) = A_{i-}^T (A_{i-}x - b_i).$$
(2.19)

Die Kosten einer Iteration liegen beim stochastischen Gradientenverfahren also in $\mathcal{O}(p)$ im Vergleich zu $\mathcal{O}(pn)$ beim traditionellen Gradientenverfahren.

2.4.2 Konvergenzanalyse

Nachdem wir bereits festgestellt haben, dass die Anzahl der Operationen eines Iterationsschrittes konstant bezüglich der Anzahl der Teilfunktionen n ist, verbleibt nun noch die Analyse der Konvergenzgeschwindigkeit. Da die Iterierten bei den stochastischen Varianten des Gradientenverfahrens von Zufallsvariablen abhängen, wird bei deren Konvergenzanalyse von Liu (2015) und Roux et al. (2012) der Erwartungswert des Fehlers $\mathbb{E}||x^k - x^*||^2$ betrachtet. Der nachfolgende Satz von Liu (2015) zeigt, dass die Approximation des Gradienten $\nabla g(x^k)$ mittels des stochastischen Gradienten $\nabla f_{i_k}(x^k)$ zur sublinearen Konvergenz des Verfahrens führt.

Proposition 10. (Sublineare Konvergenz des stochastischen Gradientenverfahrens (Liu, 2015)) Sei G > 0 eine Konstante, die die Ungleichung $\mathbb{E} \|\nabla f_{i_k}(x^k)\|^2 \leq G^2$ erfüllt, so konvergiert das stochastische Gradientenverfahren für die fallende Schrittweite $\alpha_k = \frac{1}{\mu k}$ Q-sublinear und es gilt

$$\mathbb{E}\|x^{k} - x^{*}\|^{2} \le \frac{\max\left(\|x^{1} - x^{*}\|^{2}, \frac{G^{2}}{\mu^{2}}\right)}{k}.$$
(2.20)

Beweis. Aufgrund der starken Konvexität von g gilt

$$(\nabla g(x^k) - \nabla g(x^*))^T (x^k - x^*) = \nabla g(x^k)^T (x^k - x^*) \ge \mu \|x^k - x^*\|^2.$$
(2.21)

Des Weiteren gilt für den Erwartungswert des Fehlers der (k+1)-ten Iterierten

$$\mathbb{E} \|x^{k+1} - x^*\|^2 = \mathbb{E} \|x^k - \alpha_k \nabla f_{i_k}(x^k) - x^*\|^2$$

= $\mathbb{E} \|x^k - x^*\|^2 - 2\alpha_k \mathbb{E} \left[\nabla f_{i_k}(x^k)^T (x^k - x^*) \right] + \alpha_k^2 \mathbb{E} \|\nabla f_{i_k}(x^k)\|^2$
 $\leq \mathbb{E} \|x^k - x^*\|^2 - 2\alpha_k \mathbb{E} \left[\nabla g(x^k)^T (x^k - x^*) \right] + \alpha_k^2 G^2.$ (2.22)

Ungleichung (2.21) angewandt auf Ungleichung (2.22) ergibt

$$\mathbb{E} \|x^{k+1} - x^*\|^2 \le \mathbb{E} \|x^k - x^*\|^2 - 2\alpha_k \mu \mathbb{E} \|x^k - x^*\|^2 + \alpha_k^2 G^2$$

= $(1 - 2\mu\alpha_k) \mathbb{E} \|x^k - x^*\|^2 + \alpha_k^2 G^2.$ (2.23)

Daraus erhalten wir das gewünschte Resultat (Ungleichung (2.20)) mithilfe eines Induktionsbeweises.

Induktionshypothese: Ungleichung (2.20) gilt.

Induktionsstart: Trivialerweise gilt die Ungleichung

$$\mathbb{E}\|x^{1} - x^{*}\|^{2} \le \frac{\max\left(\|x^{1} - x^{*}\|^{2}, \frac{G^{2}}{\mu^{2}}\right)}{1}$$
(2.24)

aufgrund der Definition des Maximums.

Induktionsschritt: Unter der Annahme, die Induktionshypothese gelte für k, zeigen wir, dass sie auch für k + 1 gilt. Dabei verwenden wir die Kurzschreibweise $M := \max\left(\|x^1 - x^*\|^2, \frac{G^2}{\mu^2}\right)$. Aus Ungleichung (2.23) erhalten wir mithilfe der Induktionshypothese und der Definition des Maximums das gewünschte Ergebnis

$$\mathbb{E} \|x^{k+1} - x^*\|^2 \leq \left(1 - \frac{2}{k}\right) \mathbb{E} \|x^k - x^*\|^2 + \frac{1}{\mu^2 k^2} G^2 \quad \text{[IH]} \\
\leq \left(1 - \frac{2}{k}\right) \frac{M}{k} + \frac{1}{\mu^2 k^2} G^2 \quad \text{[Def. Maximum]} \\
\leq \left(\frac{1}{k} - \frac{2}{k^2}\right) M + \frac{M}{k^2} \\
= \left(\frac{1}{k} - \frac{1}{k^2}\right) M \\
\leq \frac{M}{k+1}.$$
(2.25)

Dieses Ergebnis ist intuitiv sinnvoll: Durch das Ersetzen des Gradienten mit einer Approximation scheinen wir an Konvergenzgeschwindigkeit zu verlieren. Die Folge der Iterierten konvergiert nicht mehr Q-linear sondern scheinbar nur noch Q-sublinear.

Unter gewissen Vorraussetzungen an die zu minimierende Zielfunktion zeigen Nemirovski and Udin (1983), Nemirovski et al. (2009) und Agarwal et al. (2012), dass die optimale Konvergenzrate für den Fehler der Zielfunktionen $\mathbb{E}[g(x^k)] - g(x^*)$ bei der stark konvexen Optimierung mittels Algorithmen, die nur durch unverzerrte Messungen der Zielfunktion auf die Funktion zugreifen können, Q-sublinar durch

$$\mathbb{E}[g(x^k)] - g(x^*) \in \mathcal{O}\left(\frac{1}{k}\right)$$
(2.26)

2.4. STOCHASTISCHES GRADIENTENVERFAHREN

gegeben ist. Aus der Abschätzung

$$\frac{\mu}{2} \|x^k - x^*\|^2 \le g(x^k) - g(x^*) \le \frac{L}{2} \|x^k - x^*\|^2,$$
(2.27)

die aufgrund der starken Konvexität von g und der Lipschitz Stetigkeit von ∇g gilt, folgt die analoge Aussage für den quadrierten Fehler der Iterierten. Somit konvergiert die Folge der Iterierten beim stochastischen Gradientenverfahren langsamer als die des herkömmlichen Gradientenverfahrens, welche Q-linear konvergiert. Im Gegensatz zum Gradientenverfahren, welches Q-linear konvergiert, konvergiert das stochastische Gradientenverfahren nur Q-sublinear.

Die Annahme $\mathbb{E} \|\nabla f_{i_k}(x^k)\|^2 \leq G^2$ mit G > 0 aus Proposition 10 ist insbesondere dann erfüllt, wenn die Varianz der Komponenten des stochastischen Gradienten beschränkt ist.

Kapitel 3

Das Stochastic-Average-Gradient Verfahren

Das von Roux et al. (2012) eingeführte Stochastic-Average-Gradient Verfahren (SAG-Verfahren) ist eine besonders interessante Erweiterung des stochastischen Gradientenverfahrens, weil sie die Konvergenzgeschwindigkeit des Verfahrens unter der Zusatzannahme, dass die Funktionen von einem endlichen Datensatz stammen, erhöht. Im Gegensatz zum stochastischem Gradientenverfahren, welches sublinear konvergiert, konvergiert das SAG-Verfahren linear.

Das SAG Verfahren ist durch die Rekursion

$$x^{0} \in \mathbb{R}^{p}$$

$$x^{k+1} = x^{k} - \frac{\alpha_{k}}{n} \sum_{i=1}^{n} y_{i}^{k+1}$$
(3.1)

gegeben, wobei

$$y_i^{k+1} \coloneqq \begin{cases} \nabla f_i(x^k) & \text{wenn } i = i_k, \\ y_i^k & \text{sonst,} \end{cases}$$
(3.2)

und i_k gleichverteilt aus $\{1, \ldots, n\}$ gewählt wird. Anstatt wie im traditionellen stochastischen Gradientenverfahren die in den vorherigen Schritten berechneten Teilgradienten zu verwerfen, werden deren aktuellsten Versionen in y^k mitgeführt. Die Verbesserung der Konvergenzgeschwindigkeit erfordert also zusätzlichen Speicherplatz. Im einfachsten Fall wird $y_i^0 = 0$ für i in $\{1, \ldots, n\}$ gewählt, aber es sind auch andere Initialisierungen für den Speicher der Teilgradienten möglich.

3.1 Schrittkosten

Wie beim stochastischen Gradientenverfahren hängen die Schrittkosten hauptsächlich von der Berechnung des Gradienten der Teilfunktion $\nabla f_i(x^k)$ ab, zusätzlich muss in jedem Schritt die Summe $\sum_{i=1}^n y_i^{k+1}$ gebildet werden.

Beim Beispiel des Lösens von überbestimmten Gleichungssystemen, zur Erinnerung:

$$g(x) = \frac{1}{2n} \|Ax - b\|^{2}$$

= $\frac{1}{2n} \sum_{i=1}^{n} |A_{i-}^{T}x - b_{i}|^{2}, A \in \mathbb{R}^{n \times p}$
 $\nabla f_{i}(x) = A_{i-}^{T}(A_{i-}x - b_{i}),$ (3.3)

und

ergibt das einen Rechenaufwand in $\mathcal{O}(n+p)$, *n* Operationen für die Summe $\sum_{i=1}^{n} y_i^k$ und *p* für die Berechnung des Gradienten der Teilfunktion. Im Vergleich dazu liegt der Rechenaufwand beim traditionelle Gradientenverfahren $\mathcal{O}(np)$ und beim stochastische Gradientenverfahren $\mathcal{O}(p)$. Der Aufwand der Iterationen des SAG-Verfahren ist zwischen dem der beiden anderen Verfahren. Allerdings liegt der Aufwand des SAG-Verfahrens für praxisrelevante Dimensionen von *n* und *p* deutlich näher bei dem des stochastischen Gradientenverfahren.

3.2 Theoretische Konvergenzanalyse

Folgender Satz zeigt die bereits erwähnte Q-lineare Konvergenz des SAG-Verfahrens.

Proposition 11. (Konvergenzrate SAG-Verfahren) Mit der konstanten Schrittweite $\alpha_k = \frac{1}{2nL}$ konvergiert das SAG-Verfahren exponentiell und die durch Gleichung (3.1) und Gleichung (3.2) gegebenen Iterierten erfüllen für k > 1 die Eigenschaft:

$$\mathbb{E}\left[\|x^{k} - x^{*}\|^{2}\right] \leq \left(1 - \frac{\mu}{8Ln}\right)^{k} \left[3\|x_{0} - x^{*}\|^{2} + \frac{9\sigma^{2}}{4L^{2}}\right],\tag{3.4}$$

wobei n die Anzahl der Teilfunktionen, L eine Lipschitzkonstante der Teilgradienten, μ die zur stark konvexen Funktion g gehörige Konstante und $\sigma^2 \coloneqq \frac{1}{n} \sum_{i=1}^n \|f_i(x^*)\|^2$ ist.

Beweis. Im Folgenden werden wir den Beweis der Konvergenzrate von Roux et al. (2012) nachvollziehen.

3.2.1 Problemformulierung und Notation

Wir betrachten die μ -stark konvexe Funktion $g = \frac{1}{n} \sum_{i=1}^{n} f_i$, wobei die Funktionen f_1, \ldots, f_n konvexe Funktionen von \mathbb{R}^p nach \mathbb{R} mit *L*-Lipschitz stetigen Gradienten sind. Sei x^* die eindeutige Minimalstelle von g.

Für $k \geq 1$ wird beim SAG-Verfahren die Rekursion

$$x^{k} = x^{k-1} - \frac{\alpha}{n} \sum_{i=1}^{n} y_{i}^{k},$$

wobei i_k gleichverteilt aus $\{1, \ldots, n\}$ gewählt wird und wir setzen

$$y_i^k = \begin{cases} \nabla f_i(x^{k-1}) & \text{wenn } i = i_k, \\ y_i^{k-1} & \text{sonst.} \end{cases}$$

Bezeichne z_i^k eine Zufallsvariable, die den Wert $1 - \frac{1}{n}$ mit Wahrscheinlichkeit $\frac{1}{n}$ annimmt und den Wert $-\frac{1}{n}$ mit Wahrscheinlichkeit $\frac{n-1}{n}$, so kann y_i^k auch als

$$y_i^k = \left(1 - \frac{1}{n}\right) y_i^{k-1} + \frac{1}{n} \nabla f_i(x^{k-1}) + z_i^k \left[\nabla f_i(x^{k-1}) - y_i^{k-1}\right]$$

ausgedrückt werden. Denn für den Fall z_i^k = 1 – $\frac{1}{n}$ gilt die Gleichung

$$y_i^k = \left(1 - \frac{1}{n}\right) y_i^{k-1} + \frac{1}{n} \nabla f_i(x^{k-1}) + \left(1 - \frac{1}{n}\right) \left[\nabla f_i(x^{k-1}) - y_i^{k-1}\right]$$
$$= \nabla f_i(x^{k-1})$$

und für den Fall $z_i^k = -\frac{1}{n}$ gilt die Gleichung

$$y_i^k = \left(1 - \frac{1}{n}\right) y_i^{k-1} + \frac{1}{n} \nabla f_i(x^{k-1}) - \frac{1}{n} \left[\nabla f_i(x^{k-1}) - y_i^{k-1}\right]$$

= y_i^{k-1} .

Setzt man diese Darstellung in die Definition von x^k ein, so erhält man

$$x^{k} = x^{k-1} - \frac{\alpha}{n} \sum_{i=1}^{n} \left[\left(1 - \frac{1}{n} \right) y_{i}^{k-1} + \frac{1}{n} \nabla f_{i}(x^{k-1}) + z_{i}^{k} \left[\nabla f_{i}(x^{k-1}) - y_{i}^{k-1} \right] \right].$$

Aufgrund der Zugehörigkeit von y_i^k und $\nabla f_i(x^{k-1})$ zu \mathbb{R}^p und der Definition von ∇g kann x^k als

$$x^{k} = x^{k-1} - \frac{\alpha}{n} \sum_{i=1}^{n} \left[\left(1 - \frac{1}{n} \right) I y_{i}^{k-1} + z_{i}^{k} I \left[\nabla f_{i}(x^{k-1}) - y_{i}^{k-1} \right] \right] - \frac{\alpha}{n} \nabla g(x^{k-1})$$

geschrieben werden, wobe
iIdie Einheitsmatrix in $\mathbb{R}^{p\times p}$ bezeichnet. Führt man zusätzlich die Matrizen

$$e = \begin{pmatrix} I \\ \vdots \\ I \end{pmatrix} \in \mathbb{R}^{np \times p}, \ \nabla f(x) = \begin{pmatrix} \nabla f_1(x) \\ \vdots \\ \nabla f_n(x) \end{pmatrix} \in \mathbb{R}^{np}, \quad z^k = \begin{pmatrix} z_1^k I \\ \vdots \\ z_n^k I \end{pmatrix} \in \mathbb{R}^{np \times p}$$

ein, so erhält man eine Matrixschreibweise für x^k , nämlich,

$$x^{k} = x^{k-1} - \frac{\alpha}{n} \left[\left(1 - \frac{1}{n} \right) e^{T} y^{k-1} + \nabla g(x^{k-1}) + (z^{k})^{T} \left[\nabla f(x^{k-1}) - y^{k-1} \right] \right].$$

Mit obiger Definition von z^k gilt $\mathbb{E}[z^k(z^k)^T] = \frac{1}{n}I - \frac{1}{n^2}ee^T$. Außerdem sind die Variablen z_1^k, \ldots, z_n^k für gegebenes k nicht unabhängig. Die von den Zufallsvariablen z^1, \ldots, z^k induzierte σ -Algebra wird als \mathcal{F}_k bezeichnet.

Des Weiteren nutzen wir die Notation

$$\theta^{k} = \begin{pmatrix} y_{1}^{k} \\ \vdots \\ y_{n}^{k} \\ x^{k} \end{pmatrix} \in \mathbb{R}^{(n+1)p}, \ \theta^{*} = \begin{pmatrix} \nabla f_{1}(x^{*}) \\ \vdots \\ \nabla f_{n}(x^{*}) \\ x^{*} \end{pmatrix} \in \mathbb{R}^{(n+1)p}$$

und

$$\sigma^{2} = \frac{1}{n} \sum_{i=1}^{n} \|\nabla f_{i}(x^{*})\|^{2}.$$

Ist M eine $tp \times tp$ Matrix und m eine $tp \times p$ Matrix, dann bezeichne

- diag(M) die $tp \times p$ Matrix, die die Konkatenierung der $t(p \times p)$ -Blöcke auf der Diagonalen von M enthält und
- Diag(m) die $tp \times tp$ Matrix, deren $(p \times p)$ -Blöcke auf der Diagonale den $(p \times p)$ -Blöcken aus m entsprechen.

3.2.2 Beweis nach Roux et al. (2012)

Beim Beweis wird wie folgt vorgegangen:

1. Zunächst wird mithilfe einer sogenannten Lyapunov Funktion Q von $\mathbb{R}^{(n+1)p}$ nach \mathbb{R} eine lineare konvergente Folge $(\mathbb{E}Q(\theta^k))_{k\in\mathbb{N}}$ konstruiert.

2. Anschließend wird gezeigt, dass die k-te Differenz $||x^k - x^*||^2$ vom k-ten Folgenglied $Q(\theta^k)$ um eine Konstante dominiert wird.

Definition 12. (Lyapunov Funktion nach Joseph P. and Solomon (1967)) Sei $q \in \mathbb{N}$ und $\mathcal{D} \subseteq \mathbb{R}^{q}$ offen. Eine Funktion $V : \mathcal{D} \to \mathbb{R}$ heißt Lyapunov Funktion, wenn sie stetig differenzierbar und positiv definit, d.h. V(y) > 0 für $y \neq 0$, ist.

1. Lineare Konvergenz der Lyapunov Funktion

Für die konstante Schrittweite $\alpha = \frac{1}{2nL}$ betrachten wir die quadratische Lyapunov Funktion

$$Q(\theta^k) = (\theta^k - \theta^*)^T \begin{pmatrix} A & b \\ b^T & c \end{pmatrix} (\theta^k - \theta^*),$$
(3.5)

 mit

$$A = 3n\alpha^{2}I + \frac{\alpha^{2}}{n}\left(\frac{1}{n} - 2\right)ee^{T}$$

$$b = -\alpha\left(1 - \frac{1}{n}\right)e$$

$$c = I$$

$$S = 3n\alpha^{2}I$$

$$b - \frac{\alpha}{n}ec = -\alpha e.$$
(3.6)

Zur Erinnerung θ^k und θ^* sind durch

$$\theta^{k} = \begin{pmatrix} y_{1}^{k} \\ \vdots \\ y_{n}^{k} \\ x^{k} \end{pmatrix} \in \mathbb{R}^{(n+1)p} \quad \text{und} \quad \theta^{*} = \begin{pmatrix} \nabla f_{1}(x^{*}) \\ \vdots \\ \nabla f_{n}(x^{*}) \\ x^{*} \end{pmatrix} \in \mathbb{R}^{(n+1)p}$$

gegeben

Um die lineare Konvergenz der resultierenden Folge $(\mathbb{E}Q(\theta^k))_{k\in\mathbb{N}}$ zu zeigen, muss ein $\delta > 0$ gefunden werden für welches die Ungleichung $\mathbb{E}Q(\theta^k) \leq (1-\delta)\mathbb{E}Q(\theta^{k-1})$ gilt. Dazu wird das nachfolgende Lemma, welches eine alternative Darstellung für den Ausdruck

$$\mathbb{E}\left[(\theta^k - \theta^*)^T \begin{pmatrix} A & b \\ b^T & c \end{pmatrix} (\theta^k - \theta^*) |\mathcal{F}_{k-1} \right]$$

liefert verwendet. Dabei bezeichnet $\mathbb{E}[X|\mathcal{F}_{k-1}]$ den bedingten Erwartungswert der Zufallsvariable X gegeben der Unter- σ -Algebra \mathcal{F}_{k-1} . Die σ -Algebra \mathcal{F}_{k-1} ist eine Unter- σ -Algebra von \mathcal{F}_k . **Lemma 13.** Für eine Matrix $P = \begin{pmatrix} A & b \\ b^T & c \end{pmatrix}$, mit $A \in \mathbb{R}^{np \times np}$, $B \in \mathbb{R}^{np \times p}$ und $c \in \mathbb{R}^{p \times p}$, gilt

$$\mathbb{E}\left[\left(\theta^{k}-\theta^{*}\right)^{T}\begin{pmatrix}A & b\\b^{T} & c\end{pmatrix}\left(\theta^{k}-\theta^{*}\right)|\mathcal{F}_{k-1}\right] \\
=\left(y^{k-1}-\nabla f(x^{*})\right)^{T}\left[\left(1-\frac{2}{n}\right)S+\frac{1}{n}\operatorname{Diag}(\operatorname{diag}(S))\right]\left(y^{k-1}-\nabla f(x^{*})\right) \\
+\frac{1}{n}(\nabla f(x^{k-1})-\nabla f(x^{*}))^{T}\operatorname{Diag}(\operatorname{diag}(S))(\nabla f(x^{k-1})-\nabla f(x^{*})) \\
+\frac{2}{n}(y^{k-1}-\nabla f(x^{*}))^{T}[S-\operatorname{Diag}(\operatorname{diag}(S))](\nabla f(x^{k-1})-\nabla f(x^{*})) \\
+2\left(1-\frac{1}{n}\right)(y^{k-1}-\nabla f(x^{*}))^{T}\left[b-\frac{\alpha}{n}ec\right](x^{k-1}-x^{*}) \\
+\frac{2}{n}(\nabla f(x^{k-1})-\nabla f(x^{*}))^{T}\left[b-\frac{\alpha}{n}ec\right](x^{k-1}-x^{*}) \\
+(x^{k-1}-x^{*})^{T}c(x^{k-1}-x^{*}),$$
(3.7)

mit

$$S = A - \frac{\alpha}{n}be^T - \frac{\alpha}{n}eb^T + \frac{\alpha^2}{n^2}ece^T.$$
(3.8)

Beweis. Der interessierte Leser findet den Beweis für dieses Lemma im Anhang der Publikation von Roux et al. (2012). $\hfill \Box$

Mithilfe des obigen Lemmas erhalten wir

$$\mathbb{E}\left[Q(\theta^{k})|\mathcal{F}_{k-1}\right] = \mathbb{E}\left[\left(\theta^{k} - \theta^{*}\right)^{T} \begin{pmatrix} A & b \\ b^{T} & c \end{pmatrix} (\theta^{k} - \theta^{*})|\mathcal{F}_{k-1}\right] \\
= (y^{k-1} - \nabla f(x^{*}))^{T}\left[\left(1 - \frac{2}{n}\right)S + \frac{1}{n}\operatorname{Diag}(\operatorname{diag}(S))\right](y^{k-1} - \nabla f(x^{*})) \\
+ \frac{1}{n}(\nabla f(x^{k-1}) - \nabla f(x^{*}))^{T}\operatorname{Diag}(\operatorname{diag}(S))(\nabla f(x^{k-1}) - \nabla f(x^{*})) \\
+ \frac{2}{n}(y^{k-1} - \nabla f(x^{*}))^{T}[S - \operatorname{Diag}(\operatorname{diag}(S))](\nabla f(x^{k-1}) - \nabla f(x^{*})) \\
+ 2\left(1 - \frac{1}{n}\right)(y^{k-1} - \nabla f(x^{*}))^{T}\left[b - \frac{\alpha}{n}ec\right](x^{k-1} - x^{*}) \\
+ \frac{2}{n}(\nabla f(x^{k-1}) - \nabla f(x^{*}))^{T}\left[b - \frac{\alpha}{n}ec\right](x^{k-1} - x^{*}) \\
+ (x^{k-1} - x^{*})^{T}c(x^{k-1} - x^{*}).$$
(3.9)

Für unsere Wahl von A, b und c gelten die Gleichungen

$$S - \text{Diag}(\text{diag}(S)) = 3n\alpha^2 I - 3n\alpha^2 I = 0$$
(3.10)

und aufgrund der Konvexität von g

$$e^{T}(\nabla f(x^{k-1}) - \nabla f(x^{*})) = n(\nabla g(x^{k-1}) - \underbrace{\nabla g(x^{*})}_{=0}) = n \nabla g(x^{k-1}).$$
(3.11)

Einsetzen von Gleichung (3.10) in Gleichung (3.9) liefert

$$\mathbb{E}\left[Q(\theta^{k})|\mathcal{F}_{k-1}\right] = \left(1 - \frac{1}{n}\right) 3n\alpha^{2}(y^{k-1} - \nabla f(x^{*}))^{T}(y^{k-1} - \nabla f(x^{*})) + (x^{k-1} - x^{*})^{T}(x^{k-1} - x^{*}) - \frac{2\alpha}{n}(x^{k-1} - x^{*})^{T}e^{T}(\nabla f(x^{k-1}) - \nabla f(x^{*})) + 3\alpha^{2}(\nabla f(x^{k-1}) - \nabla f(x^{*}))^{T}(\nabla f(x^{k-1}) - \nabla f(x^{*})) - 2\alpha\left(1 - \frac{1}{n}\right)(y^{k-1} - \nabla f(x^{*}))^{T}(x^{k-1} - x^{*})$$
(3.12)

und darauffolgendes einsetzen von Gleichung (3.11) liefert

$$\mathbb{E}\left[Q(\theta^{k})|\mathcal{F}_{k-1}\right] = \left(1 - \frac{1}{n}\right) 3n\alpha^{2}(y^{k-1} - \nabla f(x^{*}))^{T}(y^{k-1} - \nabla f(x^{*})) + (x^{k-1} - x^{*})^{T}(x^{k-1} - x^{*}) - 2\alpha(x^{k-1} - x^{*})^{T}\nabla g(x^{k-1}) + 3\alpha^{2}(\nabla f(x^{k-1}) - \nabla f(x^{*}))^{T}(\nabla f(x^{k-1}) - \nabla f(x^{*})) - 2\alpha\left(1 - \frac{1}{n}\right)(y^{k-1} - \nabla f(x^{*}))^{T}(x^{k-1} - x^{*}).$$
(3.13)

Nähere Betrachtung der dritten Zeile von Gleichung (3.13) liefert

$$(\nabla f(x^{k-1}) - \nabla f(x^*))^T (\nabla f(x^{k-1}) - \nabla f(x^*)) = \sum_{i=1}^n \|\nabla f_i(x^{k-1}) - \nabla f_i(x^*)\|^2$$

$$\leq \sum_{i=1}^n L (\nabla f_i(x^{k-1}) - \nabla f_i(x^*))^T (x^{k-1} - x^*) \quad (3.14)$$

$$= nL (\nabla g(x^{k-1}) - \nabla g(x^*))^T (x^{k-1} - x^*)$$

$$= nL \nabla g(x^{k-1})^T (x^{k-1} - x^*),$$

wobei die Ungleichung in der zweiten Zeile von (Nesterov, 2004, Theorem 2.1.5) stammt. Einsetzen von Gleichung (3.14) in Gleichung (3.13) liefert

$$\mathbb{E}\left[Q(\theta^{k})|\mathcal{F}_{k-1}\right] = \left(1 - \frac{1}{n}\right) 3n\alpha^{2}(y^{k-1} - \nabla f(x^{*}))^{T}(y^{k-1} - \nabla f(x^{*})) + (x^{k-1} - x^{*})^{T}(x^{k-1} - x^{*}) - 2\alpha(x^{k-1} - x^{*})^{T}\nabla g(x^{k-1}) + 3\alpha^{2}nL(x^{k-1} - x^{*})^{T}\nabla g(x^{k-1}) - 2\alpha\left(1 - \frac{1}{n}\right)(y^{k-1} - \nabla f(x^{*}))^{T}(x^{k-1} - x^{*}).$$
(3.15)

Des Weiteren gilt per Definition

$$(1-\delta)Q(\theta^{k-1}) = (1-\delta)(\theta^{k-1} - \theta^*)^T \begin{pmatrix} A & b \\ b^T & c \end{pmatrix} (\theta^{k-1} - \theta^*)$$

$$= (1-\delta)(y^{k-1} - \nabla f(x^*))^T \left[3n\alpha^2 I + \frac{\alpha^2}{n} \left(\frac{1}{n} - 2 \right) ee^T \right] (y^{k-1} - \nabla f(x^*))$$
(3.16)
$$+ (1-\delta)(x^{k-1} - x^*)^T (x^{k-1} - x^*)$$

$$- 2\alpha (1-\delta) \left(1 - \frac{1}{n} \right) (y^{k-1} - \nabla f(x^*))^T e(x^{k-1} - x^*).$$

Für die Differenz $\mathbb{E}[Q(\theta^k)|\mathcal{F}_{k-1}] - (1-\delta)Q(\theta^{k-1})$ gilt die Abschätzung

$$\mathbb{E}[Q(\theta^{k})|\mathcal{F}_{k-1}] - (1-\delta)Q(\theta^{k-1}) \\
\leq (y^{k-1} - \nabla f(x^{*}))^{T} \left[3n\alpha^{2} \left(\delta - \frac{1}{n} \right) I + (1-\delta) \frac{\alpha^{2}}{n} \left(2 - \frac{1}{n} \right) ee^{T} \right] (y^{k-1} - \nabla f(x^{*})) \\
+ \delta (x^{k-1} - x^{*})^{T} (x^{k-1} - x^{*}) \\
- (2\alpha - 3\alpha^{2}nL)(x^{k-1} - x^{*})^{T} \nabla g(x^{k-1}) \\
- 2\alpha\delta \left(1 - \frac{1}{n} \right) (y^{k-1} - \nabla f(x^{*}))^{T} e(x^{k-1} - x^{*}).$$
(3.17)

Diese Abschätzung kann mittels Eigenschaften negativ definiter Matrizen weiter verfeinert werden. Für eine negativ definite Matrix M und zwei Vektoren s und t gilt nämlich

$$(s + \frac{1}{2}M^{-1}t)^{T}M(s + \frac{1}{2}M^{-1}t) \le 0, \qquad (3.18)$$

und deshalb insbesondere

$$s^{T}Ms + s^{T}t \le -\frac{1}{4}t^{T}M^{-1}t.$$
(3.19)

Die Verwendung von Ungleichung (3.19) mit

$$M = \left[3n\alpha^2 \left(\delta - \frac{1}{n}\right)I + (1 - \delta)\frac{\alpha^2}{n} \left(2 - \frac{1}{n}\right)ee^T\right]$$
$$= \left[3n\alpha^2 \left(\delta - \frac{1}{n}\right)\left(I - \frac{ee^T}{n}\right) + \alpha \left(3n\delta - 1 - 2\delta + \frac{\delta - 1}{n}\right)\frac{ee^T}{n}\right]$$
$$s = y^{k-1} - \nabla f(x^*)$$
$$t = -2\alpha\delta(1 - \frac{1}{n})e(x^{k-1} - x^*)$$

liefert

$$(y^{k-1} - \nabla f(x^*))^T \left[3n\alpha^2 \left(\delta - \frac{1}{n} \right) I + (1 - \delta) \frac{\alpha^2}{n} \left(2 - \frac{1}{n} \right) ee^T \right] (y^{k-1} - \nabla f(x^*)) - 2\alpha \delta \left(1 - \frac{1}{n} \right) (y^{k-1} - \nabla f(x^*))^T e(x^{k-1} - x^*) \leq -\alpha^2 \delta^2 \left(1 - \frac{1}{n} \right)^2 (x^{k-1} - x^*)^T e^T \left[3n\alpha^2 \left(\delta - \frac{1}{n} \right) \left(I - \frac{ee^T}{n} \right) \right] + \alpha \left(3n\delta - 1 - 2\delta + \frac{\delta - 1}{n} \right) \frac{ee^T}{n} \right]^{-1} e(x^{k-1} - x^*)$$
(3.20)
$$= \frac{\alpha^2 \delta^2 \left(1 - \frac{1}{n} \right)^2 n}{\alpha^2 [3n\delta - 1 - 2\delta + \frac{\delta - 1}{n}]} \| x^{k-1} - x^* \|^2 = \frac{\delta^2 \left(1 - \frac{1}{n} \right)^2 n}{3n\delta - 1 - 2\delta + \frac{\delta - 1}{n}} \| x^{k-1} - x^* \|^2,$$

wobei anzumerken ist, dass M für hinreichend kleine $\delta \leq \frac{1}{3n}$ negativ definit ist. Somit gilt Ungleichung (3.20) für $\delta \leq \frac{1}{3n}$ und es folgt die neue Schranke

$$\mathbb{E}[Q(\theta^{k})|\mathcal{F}_{k-1}] - (1-\delta)Q(\theta^{k-1}) \leq -(2\alpha - 3\alpha^{2}nL)(x^{k-1} - x^{*})^{T} \nabla g(x^{k-1}) \\ + \left(\delta - \frac{\delta^{2}(1-\frac{1}{n})^{2}n}{3n\delta - 1 - 2\delta + \frac{\delta - 1}{n}}\right) \|x^{k-1} - x^{*}\|^{2}.$$
(3.21)

An dieser Stelle wird die starke Konvexität von g verwendet. Aus der starken Konvexität von g folgt nämlich unmittelbar die Ungleichung

$$\|x^{k-1} - x^*\|^2 \le \frac{1}{\mu} (x^{k-1} - x^*)^T \nabla g(x^{k-1}).$$
(3.22)

Mithilfe von Ungleichung (3.22) erhalten wir die Abschätzung

$$\mathbb{E}[Q(\theta^{k})|\mathcal{F}_{k-1}] - (1-\delta)Q(\theta^{k-1}) \leq -\left(2\alpha - 3\alpha^{2}nL + \frac{\delta^{2}\left(1-\frac{1}{n}\right)^{2}}{3n\delta - 1 - 2\delta + \frac{\delta - 1}{n}\frac{n}{\mu}} - \frac{\delta}{\mu}\right)$$
(3.23)
$$(x^{k-1} - x^{*})^{T} \nabla g(x^{k-1}).$$

Um die lineare Konvergenz der Folge $\left(\mathbb{E}Q(\theta^k)\right)_{k\in\mathbb{N}}$, d.h.

$$\mathbb{E}Q(\theta^k) \le (1-\delta)\mathbb{E}Q(\theta^{k-1}),$$

zu beweisen, zeigen wir, dass die rechte Seite von Abschätzung (3.23) negativ ist. Aufgrund von Ungleichung (3.22) wissen wir, dass $(x^{k-1} - x^*)^T \nabla g(x^{k-1})$ positiv ist. Daher muss nur noch die Positivität von $\left(2\alpha - 3\alpha^2 nL + \frac{\delta^2(1-\frac{1}{n})^2}{3n\delta - 1-2\delta + \frac{\delta^{-1}}{n}}\frac{n}{\mu} - \frac{\delta}{\mu}\right)$ gezeigt werden. Mit der Wahl $\delta = \frac{\mu}{8nL}$ und der Schrittweite $\alpha = \frac{1}{2nL}$ erhält man

$$2\alpha - 3\alpha^{2}nL + \frac{\delta^{2}\left(1 - \frac{1}{n}\right)^{2}}{3n\delta - 1 - 2\delta + \frac{\delta - 1}{n}}\frac{n}{\mu} - \frac{\delta}{\mu} = \frac{1}{nL} - \frac{3}{4nL} - \frac{1}{8nL} - \frac{\delta^{2}(1 - \frac{1}{n})^{2}\frac{n}{\mu}}{1 - 3n\delta + 2\delta + \frac{1 - \delta}{n}}$$

$$\geq \frac{1}{8nL} - \frac{\frac{\delta^{2}n}{\mu}}{1 - 3n\delta}$$

$$= \frac{1}{8nL} - \frac{\frac{\omega}{64nL^{2}}}{1 - \frac{3\mu}{8L}}$$

$$\geq \frac{1}{8nL} - \frac{\frac{\omega}{64nL^{2}}}{1 - \frac{3}{8}}$$

$$= \frac{1}{8nL} - \frac{\frac{\mu}{40nL^{2}}}{1 - \frac{3}{8}}$$

$$= \frac{1}{8nL} - \frac{1}{40nL}$$

$$\geq 0.$$
(3.24)

Somit gilt

$$\mathbb{E}\left[Q(\theta^k)|\mathcal{F}_{k-1}\right] - (1-\delta)Q(\theta^{k-1}) \le 0.$$

Durch die Bildung des Erwartungswertes erhalten wir das gewünschte Ergebnis

$$\mathbb{E}Q(\theta^k) - (1-\delta)\mathbb{E}Q(\theta^{k-1}) \le 0.$$

Die Folge $\left(\mathbb{E}Q(\theta^k)\right)_{k\in\mathbb{N}}$ konvergiert linear und es gilt

$$\mathbb{E}Q(\theta^k) \leq \left(1 - \frac{\mu}{8nL}\right)^k \mathbb{E}Q(\theta^0).$$

2. Dominierung von $||x^k - x^*||^2$ durch $Q(\theta^k)$

Für diesen Beweisschritt ist es hilfreich sich an die Definitionen von θ^k und θ^* zu erinnern:

$$\theta^{k} = \begin{pmatrix} y_{1}^{k} \\ \vdots \\ y_{n}^{k} \\ x^{k} \end{pmatrix} \in \mathbb{R}^{(n+1)p}, \theta^{*} = \begin{pmatrix} \nabla f_{1}(x^{*}) \\ \vdots \\ \nabla f_{n}(x^{*}) \\ x^{*} \end{pmatrix} \in \mathbb{R}^{(n+1)p}.$$

Wenn wir es schaffen zu zeigen, dass $||x^k - x^*||^2$ durch $Q(\theta^k)$ um einen Konstanten Faktor $\gamma > 0$ dominiert wird, d.h. dass die Ungleichung

$$\gamma \|x^k - x^*\|^2 \le Q(\theta^k) \tag{3.25}$$

gilt, dann folgt daraus unmittelbar die lineare Konvergenz des Verfahrens. Mithilfe der Definitionen von Q, θ^k und θ^* kann Ungleichung (3.25) folgendermaßen geschrieben werden

$$(\theta^{k} - \theta^{*})^{T} \underbrace{\begin{pmatrix} 0 & 0 \\ 0 & \gamma I \end{pmatrix}}_{:=R} (\theta^{k} - \theta^{*}) \leq (\theta^{k} - \theta^{*})^{T} \underbrace{\begin{pmatrix} A & b \\ b^{T} & c \end{pmatrix}}_{:=P} (\theta^{k} - \theta^{*}), \tag{3.26}$$

was äquivalent zur Ungleichung

$$(\theta^k - \theta^*)^T (P - R)(\theta^k - \theta^*) \ge 0 \tag{3.27}$$

ist. Offensichtlich gilt diese Ungleichung, wenn (P - R) eine positiv definite Matrix ist.

Roux et al. (2012) zeigen die positive Definitheit der Matrix (P - R) für $\gamma = \frac{1}{3}$ anhand der Schur-Komplement Bedingung.

Definition 14. (Schur-Komplement) Sei M eine $(n + m) \times (n + m)$ -Matrix, die aus vier Teilblöcken zusammengesetzt ist:

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

Dabei sei A eine $n \times n$ -, B eine $n \times m$ -, C eine $m \times n$ - und D eine $m \times m$ -Matrix. Des Weiteren sei vorausgesetzt, dass A und D invertierbar sind. Die Matrix

$$M/A = D - CA^{-1}B$$

wird als Schur-Komplement von A in M bezeichnet und die Matrix

$$M/D = A - BD^{-1}C$$

als Schur-Komplement von D in M.

Proposition 15. (Schur-Komplement Bedingung für positive Definitheit) Sei X eine symmetrische Matrix gegeben durch

$$X = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}.$$

wobei die Matrizen A und C invertierbar sind. Sei X/A das Schur-Komplement von A in X und X/C das Schur-Komplement von C in X. Dann gilt:

- X ist genau dann positiv definit, wenn A und X/A positiv definit sind.
- X ist genau dann positiv definit, wenn C und X/C positiv definit sind.

Die Matrix (P - R) ist eine symmetrische Matrix der Gestalt $\begin{pmatrix} A & b \\ b^T & c - \frac{1}{3}I \end{pmatrix}$ und A ist positiv definit. Daher muss nun nur noch die positive Definitheit des Schur-Komplements (P - R)/A nachgewiesen werden.

Das Einsetzen der Definitionen von A, b und c in die Definition von (P - R)/A liefert

$$(P-R)/A = \frac{2}{3}I - \alpha^2 \left(1 - \frac{1}{n}\right)^2 e^T \left[\left(3n\alpha^2 + \frac{\alpha^2}{n} - 2\alpha^2\right) \frac{ee^T}{n} \right]^{-1} e$$

= $\frac{2}{3}I - \frac{n\left(1 - \frac{1}{n}\right)^2}{3n + \frac{1}{n} - 2} \frac{ee^T}{n}.$ (3.28)

Es gilt

$$\frac{2}{3}I - \frac{n\left(1 - \frac{1}{n}\right)^2}{3n + \frac{1}{n} - 2} \frac{ee^T}{n} > \frac{2}{3}I - \frac{n}{3n - 2} \frac{ee^T}{n} > 0 \text{ für } n \ge 2.$$
(3.29)

Wir schreiben dabe
i $M \succ N$ für quadratische Matrizen M und
 N, wenn die MatrixM-N positiv
 definit ist.

Daraus ergibt sich die positive Definitheit von (P - R) und insbesondere

$$\mathbb{E} \|x^{k} - x^{*}\|^{2} \leq 3\mathbb{E}Q(\theta^{k}) \qquad [\text{Schritt 1}]$$

$$\leq 3\left(1 - \frac{\mu}{8nL}\right)^{k}Q(\theta^{0}), \qquad (3.30)$$

wobei

$$Q(\theta^{0}) = 3n\alpha^{2} \sum_{i=1}^{n} \|y_{i}^{0} - \nabla f_{i}(x^{*})\|^{2} + \frac{(1-2n)\alpha}{n^{2}} \|\sum_{i=1}^{n} y_{i}^{0}\|^{2}$$

$$- 2\alpha \left(1 - \frac{1}{n}\right) (x^{0} - x^{*})^{T} \left(\sum_{i=1}^{n} y_{i}^{0}\right) + \|x^{0} - x^{*}\|^{2}$$

$$= \frac{3}{4nL^{2}} \sum_{i=1}^{n} \|y_{i}^{0} - \nabla f_{i}(x^{*})\|^{2}$$

$$+ \frac{(1-2n)\alpha}{2n^{3}L} \|\sum_{i=1}^{n} y_{i}^{0}\|^{2} - \frac{n-1}{n^{2}L} (x^{0} - x^{*})^{T} \left(\sum_{i=1}^{n} y_{i}^{0}\right) + \|x^{0} - x^{*}\|^{2}$$

(3.31)

ist. Bei einer Initialisierung von y_i^0 = 0 für alle $i \in \{1, \dots, n\}$ gilt

$$Q(\theta^0) = \frac{3\sigma^2}{4L^2} + \|x^0 - x^*\|^2.$$
(3.32)

Das Einsetzen von Gleichung (3.32) in Ungleichung (3.30) liefert das gewünschte Resultat

$$\mathbb{E}\|x^{k} - x^{*}\|^{2} \leq \left(1 - \frac{\mu}{8nL}\right)^{k} \left(\frac{9\sigma^{2}}{4L^{2}} + 3\|x^{0} - x^{*}\|^{2}\right).$$
(3.33)

Kapitel 4

Numerische Ergebnisse

In diesem Kapitel wollen wir überprüfen, ob sich die vorgestellten Verfahren in einem praxisrelevantem Beispiel wirklich so verhalten, wie es die theoretische Analyse vermuten lässt. Dazu lösen wir das überbestimmte Gleichungssystem, das bei der Diskretisierung der zirkulären Radontransformation auftritt, mithilfe der verschiedenen Verfahren.

4.1 Zirkuläre Radontransformation

Die zirkuläre Radontransformation spielt bei diversen bildgebenden Verfahren in der Medizin eine große Rolle. Zum Beispiel bei der thermoakustischen und photoakustischen Tomographie. Aus physikalischer Perspektive funktionieren beide Verfahren ähnlich: Ein Körperteil wird elektromagnetischer Strahlung ausgesetzt. Dabei wird ein Teil der Strahlung vom Körperteil absorbiert, das Gewebe erwärmt sich, durch die Erwärmung dehnt sich das Gewebe aus, was zur Entstehung einer akustischen Welle, die durch den Körperteil wandert, führt. Diese akustischen Wellen werden dann von mehreren Sensoren über einen gewissen Zeitraum gemessen. Aus den Messdaten wird das Bild der Hitzeabsorbtionsfunktion des Körperteils rekonstruieren. Dies wird mithilfe der zirkulären Radontransformation g = Rf

$$g(x_0, y_0, r) = Rf(x_0, y_0, r) = \int_{\partial B((x_0, y_0), r)} f(x, y) ds,$$
(4.1)

wobei $x_0, y_0 \in \mathbb{R}$ die Koordinaten eines Kreismittelpunkts in \mathbb{R}^2 sind, r > 0 der Radius des Kreises und $f : \mathbb{R}^2 \to \mathbb{R}$ eine Bildfunktion ist, modelliert. Die Rekonstruktion des Bildes der Hitzeabsorptionsfunktion entspricht der Rekonstruktion der Bildfunktion f aus den Daten Rf entlang von Sphären verschiedener Radien, welche auf den Sensorpositionen zentriert sind. Typischerweise liegen die Sensoren auf einer Kreisbahn (Finch et al., 2007; Haltmeier, 2014).

Die Rekonstruktion erfolgt querschnittsweise, die Querschnitte werden mithilfe der Messungen aus den verschiedenen Richtungen rekonstruiert. Die Verwendung von endlich vielen Sensoren entspricht einer Diskretisierung entlang der Kreisbahn auf der die Sensoren liegen, wird zusätzlich entlang der verwendeten Radien diskretisiert, so erhält man ein Gleichungssystem

$$\tilde{g} = \tilde{R}\tilde{f},\tag{4.2}$$

wobei \tilde{g} (Abbildung 4.1 rechts) eine Matrix, die die Funktionsauswertungen von g für $a = (x_0, y_0)$ auf einer Kreisbahn und verschiedene Radien r enthält, \tilde{R} die Abbildungsmatrix der zirkulären Radontransformation und \tilde{f} (Abbildung 4.1 links) die Bildmatrix ist.



Abbildung 4.1: Schematische Darstellung der zirkulären Radontransformation. Die Sensoren sind entlang eines Kreises um das Objekt f angeordnet. Für jede Position $a = (x_0, y_0)$ auf dem Kreis und jeden Radius r liefert die zirkuläre Radontransformation Rf einen Wert Rf(a, r).

4.1.1 Aufbau und Problemformulierung

In unserem Anwendungsbeispiel rekonstruieren wir einen synthetisch erzeugten Querschnitt \hat{f} , der die Konturen eines Smileys, siehe Abbildung 4.1 links, enthält. Dessen Radontransformierte, siehe Abbildung 4.1 rechts, entspricht den Messungen von 400 zyklisch um das Objekt angeordneten Sensoren für jeweils 300 Radien, und wurde durch die Anwendung der zirkulären Radontransformation auf \tilde{f} berechnet. Die diskretisierte Radontransformierte \tilde{g} liegt also in $\mathbb{R}^{400\times 300}$.

Um ein Optimierungsproblem der gewünschten Form zu erhalten, machen wir uns die Problemstruktur zunutze und unterteilen das Gleichungssystem (4.2) in Teilgleichungssysteme; eines pro Sensor. Es gibt nämlich für jeden Sensor mehrere Gleichungen, die den Auswertungen der Radontransformation auf Sphären von verschiedenen Radien entsprechen. Somit erhalten wir die Zielfunktion

$$g(\tilde{f}) = \frac{1}{400} \sum_{a=1}^{400} \|\tilde{R}^a \tilde{f} - \tilde{g}^a\|_2^2,$$
(4.3)

wobei $a \in \{1, \ldots, 400\}$ die Diskretisierung entlang der Kreisbahn auf der die Sensoren liegen indiziert und \tilde{R}^a beziehungsweise \tilde{g}^a die zum Sensor a gehörigen Teile des Gleichungssystems (4.2), d.h. die Gleichungen für die 300 zum Sensor a gehörigen Radien, bezeichnen.

4.1.2 Landweber-Kaczmarz Verfahren

Ein etabliertes Verfahren um das Gleichungssystem (4.2) aufzulösen ist das sogenannte Landweber-Kaczmarz Verfahren (Kowar and Scherzer, 2002; Haltmeier et al., 2007), welches durch die Iterationsvorschrift

$$x^{0} \in \mathbb{R}^{p}$$

$$x^{k+1} = x^{k} - \alpha_{k} \nabla f_{j_{k}}(x^{k}) \quad \text{mit } j_{k} \coloneqq k \mod n$$
(4.4)

gegeben ist. Die Ähnlichkeit zum stochastischen Gradientenverfahren

$$x^{0} \in \mathbb{R}^{p}$$

$$x^{k+1} = x^{k} - \alpha_{k} \nabla f_{i_{k}}(x^{k}) \quad \text{mit } i_{k} \sim \mathcal{U}_{\{1,\dots,n\}}$$

$$(4.5)$$

ist dabei unschwer zu erkennen. Die beiden Verfahren unterscheiden sich lediglich in der Auswahlmethode der Teilgradienten. Im Gegensatz zum stochastischen Gradientenverfahren geschieht dies beim Landweber-Kaczmarz Verfahren auf deterministische Weise. Die Teilgradienten werden dabei der Reihe nach gewählt.

4.1.3 Numerische Experimente

Wir vergleichen die folgenden Gradienten-Methoden zur Lösung eines überbestimmten Gleichungssystems, welches bei der Zirkulären Radontransformation auftritt:

- 1. FGD: Das traditionelle Gradientenverfahren beschrieben durch Iterationsvorschrift (2.12).
- 2. LK: Das Landweber-Kaczmarz Verfahren beschrieben durch Iterationsvorschrift (4.4).
- 3. SGD: Das stochastische Gradientenverfahren beschrieben durch Iterationsvorschrift (2.16).
- 4. **SAG**: Das SAG-Verfahren beschrieben durch Iterationsvorschrift (3.1), wobei die Standardinitialisierung $y^0 = 0$ verwendet wird.
- 5. **SAG***: Das SAG-Verfahren beschrieben durch Iterationsvorschrift (3.1), wobei x^0 und y^0 mithilfe eines Zyklus des stochastischen Gradientenverfahrens initialisiert werden.

Ein Zyklus entspricht einem Schritt des traditionellen Gradientenverfahrens beziehungsweise *n* Schritten der auf Teilgradienten basierenden Verfahren. In den Experimenten wurde für alle Verfahren die konstante Schrittweite $\alpha = 1$ verwendet.

Iterationskosten

Hinsichtlich der Iterationskosten verhalten sich die Verfahren in etwa wie erwartet. Das stochastische Gradientenverfahren ist in etwa 144 mal so schnell wie das traditionelle Gradientenverfahren und das SAG-Verfahren ist nur etwas langsamer als das stochastische Gradientenverfahren. In der Theorie sollte die Iteration des stochastischen Gradientenverfahrens jedoch 400 mal so schnell sein wie die des traditionellen Verfahrens, da dort im Gegensatz zur Iteration des traditionellen Verfahrens nur einer der 400 Teilgradienten berechnet werden muss. Die Analyse der Berechnungskosten des Programmes lässt vermuten, dass diese Abweichung durch gewisse Overhead-Berechnungen in der Teilgradientenbildung zustande kommen. Eine Zusammenfassung der Iterationszeiten der verschiedenen Verfahren ist in Tabelle 4.1 zu finden.

Konvergenzrate und Genauigkeit

Abbildung 4.2 (a) stellt den Fehler der Iterierten dar, wobei die x-Achse in Zyklen unterteilt ist und die y-Achse dem integriertem Fehler der Iterierten entspricht. Abbildung 4.2 (b) stellt den Fehler der Iterierten bezüglich der Zielfunktion dar, wobei die x-Achse in Zyklen unterteilt ist und die y-Achse dem zehner-logarithmiertem Fehler der Iterierten bezüglich der Zielfunktion enthält. Ein Zyklus in Abbildung 4.2 entspricht einem Schritt des traditionellen Gradientenverfahrens beziehungsweise 400 Schritten der auf Teilgradienten basierenden Verfahren. Bei der Betrachtung von Abbildung 4.2 (a) ist zu erkennen, dass die SAG-Verfahren (SAG und SAG^{*}) in

Methode	Zeit
FGD	724 ms
LK	5 ms
SGD	5 ms
SAG	8 ms
SAG*	8 ms

Tabelle 4.1: Vergleich der Iterationskosten der verschiedenen Verfahren in Millisekunden. Die Tabelle enthält die über 50 Zyklen gemittelten Iterationszeiten der verschiedenen Verfahren.

diesem Anwendungsbeispiel nach einer oszillierenden Startphase Q-linear konvergieren. Des Weiteren konvergiert das stochastische Gradientenverfahren (SGD) wie erwartet Q-sublinear. Das Landweber-Kaczmarcz Verfahren (LK) verhält sich ähnlich zum stochastischen Gradientenverfahren also Q-sublinear. Das traditionelle Gradientenverfahren sollte theoretisch auch Q-linear konvergieren, allerdings ist dies nicht oder nur schwer zu erkennen.

Die oszillierende Startphase tritt nur bei den SAG-Verfahren auf, was vermutlich darauf zurückzuführen ist, dass durch das Speichern der aktuellsten Versionen der Gradienten vor allem anfangs, wo die Iterierten noch schlecht sind, unter anderem "falsche" Richtungen akkumuliert werden. Kann man sich nur wenige Iterationen leisten, so ist das stochastische Gradientenverfahren also zu bevorzugen. Die Initialisierung der Teilgradienten des SAG-Verfahrens durch einen Zyklus des stochastischen Gradientenverfahrens wirkt der oszillierenden Startphase kaum entgegen und führt in unserem Fall auch zu keinen sonstigen Verbesserungen. Eventuell ist eine längere Initialisierungsphase notwendig um einen Effekt zu beobachten. Dies wäre in diesem Anwendungsbeispiel durchaus möglich, da das stochastische Gradientenverfahren hier in den ersten fünf bis sechs Zyklen ohnehin besser abschneidet als das SAG-Verfahren. Es stellt sich allerdings die Frage, ob das dieses Verhalten der Regelfall ist.

Des Weiteren fällt auf, dass die auf Teilgradienten basierenden Verfahren schneller konvergieren als das traditionelle Gradientenverfahren und dass die stochastischen Verfahren schneller konvergieren als die deterministischen. Ersteres ist auf die Tatsache zurückzuführen, dass ein Zyklus des traditionellen Gradientenverfahrens einer Verbesserung der Iterierten entspricht, während ein Zyklus der auf Teilgradienten basierenden Verfahren n = 400 Verbesserungen der Iterierten entspricht. Obwohl aus der Analyse der Iterationszeiten, speziell aus der Beobachtung, dass beispielsweise eine Iteration des stochastischen Gradientenverfahrens nur 144 mal so schnell ist wie eine Iteration des traditionellen Gradientenverfahrens, hervorgeht, dass es sich hierbei um einen nicht ganz fairen Vergleich handelt, ist dieser Trend deutlich zu erkennen.

In Abbildung 4.3 werden die Verfahren hinsichtlich ihrer Genauigkeit verglichen, die x-Achse enthält dabei die zu erreichende Genauigkeit und die y-Achse die dafür notwendige Anzahl von Zyklen. Wie in Abbildung 4.2 ist auch hier zu erkennen, dass die SAG-Verfahren zwar die höchste Genauigkeit erreichen, jedoch anfangs aufgrund der Oszillationen langsamer sind als das stochastische Gradientenverfahren und als das Landweber-Kaczmarz Verfahren.

Weiters ist in Abbildung 4.3 (a) deutlich zu erkennen, dass das stochastische Gradientenverfahren in diesem Anwendungsfall besser funktioniert als das Landweber-Kaczmarz Verfahren, da es innerhalb der betrachteten Anzahl von Zyklen um eine ganze Nachkommastelle genauer ist. Diese Beobachtung ist insofern bedeuten, als das Landweber-Kaczmarz Verfahren ein gängiges Verfahren für die betrachtete Anwendungsdomäne ist.

Die Dominanz des stochastischen Gradientenverfahren gegenüber des Landweber-Kaczmarz Verfahrens spiegelt sich auch schon im Startverhalten der beiden Verfahren wieder. Dazu betrachten wir zusätzlich zu Abbildung 4.3 die Tabellen 4.2 (a) und 4.2 (b), welche die Iterierten der beiden Verfahren in Abständen von jeweils fünf Iterationen enthalten. Aus der Praxis ist bekannt, dass das Landweber-Kaczmarz Verfahren in etwa $\frac{n}{2} = 200$ Iterationen braucht um die Konturen eines Querschnittes zu rekonstruieren. Dieses Verhalten ist auch in Tabelle 4.2 zu erkennen. Im Gegensatz dazu benötigt das stochastische Gradientenverfahren nur etwa halb so viele Iterationen um Konturen in der augenscheinlich selben Qualität zu rekonstruieren.

Auch die Startschwierigkeiten des SAG-Verfahrens sind in Tabelle 4.2 (c) zu erkennen, die intensiven Gelb-töne weisen darauf hin, dass die Approximation des Gradienten im SAG-Verfahrens anfangs in eine dominante Richtung weist.



(c) SAG Verfahren

Tabelle 4.2: Illustration der Startphase der approximativen Verfahren. Zwischen den einzelnen Bildern liegen jeweils fünf Iterationen der zugehörigen Verfahren. Für jedes Verfahren werden also die Rekonstruktionen von Iteration 1 bis 196 dargestellt.



(b) Fehler bezüglich Zielfunktion

Abbildung 4.2: Vergleich der vorgestellten Methoden bezüglich der Fehler der Iterierten und der Fehler der Zielfunktion am Beispiel der zirkulären Radontransformation. Die y-Achse ist logarithmisch skaliert. Die x-Achse ist in Zyklen unterteilt. Ein Zyklus entspricht dabei einem Durchlauf durch die Daten, also einer Iteration des Gradientenverfahrens oder n Iterationen der anderen Verfahren.



(b) Genauigkeit bezüglich Zielfunktion

Abbildung 4.3: Vergleich der vorgestellten Methoden bis eine gewisse Genauigkeit erreicht ist. Die Abbildung stellt die Anzahl der Zyklen dar, die erforderlich sind um eine gewisse Genauigkeit zu erreichen.

Kapitel 5

Konklusion

In dieser Bachelorarbeit wurden stochastische Varianten des Gradientenverfahrens analysiert, die durch stochastische Näherungen des Gradienten den hohen Schrittkosten des traditionellen Gradientenverfahrens entgegenwirken sollen. Insbesondere wurden das stochastische Gradientenverfahren und das SAG-Verfahren betrachtet.

Sowohl die theoretische als auch die numerische Analyse der stochastischen Varianten des Gradientenverfahrens zeigen, dass diese im Wesentlichen ihr Ziel erreichen. Durch die Einfachheit der Approximation des Gradientens beim stochastischen Gradientenverfahren, werden die Iterationen auf Kosten der Konvergenzrate beschleunigt. Diese ist beim stochastischen Gradientenverfahren Q-sublinear. Im Gegensatz dazu ist die des Gradientenverfahrens Q-linear. Anhand einer etwas kompliziertere Approximation wird beim SAG-Verfahren eine Q-lineare Konvergenz wiederhergestellt.

Während die existierenden theoretischen Ergebnisse zu den Verfahren nahelegen, dass das traditionelle Gradientenverfahren den stochastischen Varianten gegenüber zu bevorzugen ist, wenn ausreichend Rechenleistung vorhanden ist, wirft das numerische Experiment die Frage auf wann dies tatsächlich der Fall ist. Schon bei dem kleinen Anwendungsbeispiel ist erkennbar, wie teuer die Iterationen des traditionellen Gradientenverfahrens sind. Des Weiteren erreichen beide stochastischen Verfahren in unserem numerischen Experiment hohe Genauigkeiten.

Vor allem die numerische Analyse der Verfahren verdeutlicht, dass das stochastische Gradientenverfahren den anderen betrachteten Verfahren gegenüber zu bevorzugen ist, wenn nur wenige Durchgänge durch die Teilfunktionen möglich sind. Das liegt vor allem daran, dass das SAG-Verfahren, welches ansonsten in allen Vergleichen am besten abschneidet, durch Akkumulation von "schlechten" Richtungen eine fluktuierende Startphase aufweisen kann, worum es in unserem Anwendungsbeispiel erst nach einigen Durchgängen besser ist als das stochastische Gradientenverfahren.

Obwohl wir in unserem Anwendungsbeispiel keine nennenswerte Verbesserung der Ergebnisse des SAG-Verfahrens mittels der Initialisierung der Teilgradienten durch das stochastische Gradientenverfahren beobachten konnten, liegt die Vermutung dennoch nahe, dass eine Kombination der beiden Verfahren über die problematische Startphase des SAG-Verfahrens hinweghelfen könnte, ohne an Konvergenzgeschwindigkeit zu verlieren. Daher wäre es interessant weitere Bemühungen in diese Richtung anzustellen.

In dieser Arbeit wurde nur wenig auf die konkrete Wahl der Schrittweiten der Verfahren eingegangen, da jedoch die Wahl der Schrittweiten in der Praxis eine große Rolle spielt und auch die Konvergenz der Verfahren beeinflusst, wären auch weitere Ermittlungen in diese Richtung interessant.

Literaturverzeichnis

- Alekh Agarwal, Peter L. Bartlett, Pradeep Ravikumar, and Martin J. Wainwright. Informationtheoretic lower bounds on the oracle complexity of stochastic convex optimization. *IEEE Transactions on Information Theory*, 58:3235–3249, 2012.
- David Finch, Markus Haltmeier, and Rakesh. Inversion of spherical means and the wave equation in even dimensions. *SIAM Journal on Applied Mathematics*, 68(2):392–412, 2007. doi: 10.1137/070682137. URL http://dx.doi.org/10.1137/070682137.
- Markus Haltmeier. Universal inversion formulas for recovering a function from spherical means. SIAM Journal on Mathematical Analysis, 46(1):214–232, 2014. doi: 10.1137/120881270. URL http://dx.doi.org/10.1137/120881270.
- Markus Haltmeier, Antonio Leitão, and Otmar Scherzer. Kaczmarz methods for regularizing nonlinear ill-posed equations i: convergence analysis. *Inverse Problems and Imaging*, 1(2): 289-298, 2007. ISSN 1930-8337. doi: 10.3934/ipi.2007.1.289. URL http://aimsciences. org/journals/displayArticlesnew.jsp?paperID=2247.
- LaSalle Joseph P. and Lefschetz Solomon. *Die Stabilitätstheorie von Ljapunow: die direkte Methode mit Anwendungen.* BI-Hochschultaschenbücher. Bibliographisches Institut, 1967. URL https://books.google.at/books?id=MTGXGAAACAAJ.
- Richard Kowar and Otmar Scherzer. Convergence analysis of a Landweber-Kaczmarz method for solving nonlinear ill-posed problems. Ill posed and inverse problems. 2002.
- Ji Liu. Csc 576: Stochastic gradient "descent" algorithm. Vorlesungsunterlagen, 2015. Department of Computer Sciences, University of Rochester.
- Arkadi Nemirovski and David Borisovich Udin. Problem complexity and method efficiency in optimization. Wiley-Interscience series in discrete mathematics. Wiley, Chichester, New York, 1983. ISBN 0-471-10345-4. A Wiley-Interscience publication.
- Arkadi Nemirovski, Anatoli Juditsky, Guanghui Lan, and Alexander Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4): 1574–1609, 2009. doi: 10.1137/070704277. URL http://dx.doi.org/10.1137/070704277.
- Yurii Nesterov. Introductory lectures on convex optimization : a basic course. Applied optimization. Kluwer Academic Publ., Boston, Dordrecht, London, 2004. ISBN 1-4020-7553-7.
- Rüdiger Reinhardt, Armin Hoffmann, and Tobias Gerlach. *Nichtlineare Optimierung*. Springer Spektrum, 2013.
- Nicolas Le Roux, Mark W. Schmidt, and Francis R. Bach. A stochastic gradient method with an exponential convergence rate for finite training sets. In *Neural Information Processing Systems*, 2012.