

Inhaltsverzeichnis

1	Die Idee und der Zweck des Gradientenverfahrens	2
2	Schrittweitenbestimmung mittels Gleichungssystem	2
2.1	Der Ablauf	2
2.2	Beispiel	3
3	Schrittweitenbestimmung mittels der Armijo-Bedingung	5
3.1	Der Ablauf	5
3.2	Beispiel	6

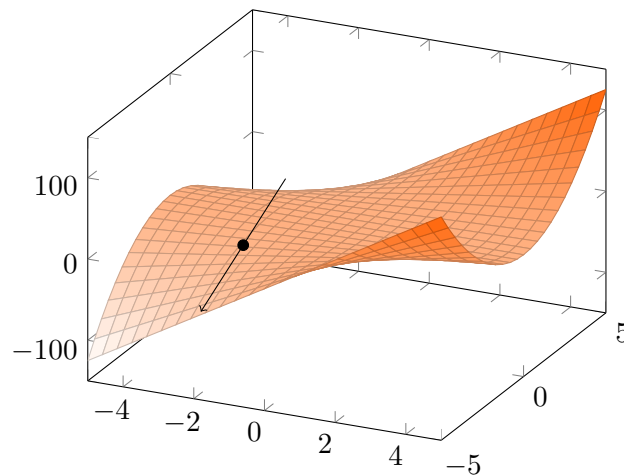


Abbildung 1: Hier zu sehen ist $-\vec{\nabla}f(-2, -2)$ der Funktion $f(x, y) = x * y^2$

1 Die Idee und der Zweck des Gradientenverfahrens

Grundsätzlich ist das Gradientenverfahren ein Optimierungsverfahren, mit welchem man das Minimum einer mehrdimensionalen Funktion errechnen kann. Der Gradient ($\nabla f(x, y)$) einer Funktion $f(x, y)$ zeigt, aus später gezeigten (HIER VERWEIS EINFÜGEN) Gründen, immer in die Richtung des steilsten Anstiegs, folglich zeigt $-\nabla f(x, y)$ in die Richtung des stärksten Abstiegs. Das Gradientenverfahren macht sich dies zunutze und man wandert eine gewisse Dauer in diese Richtung, wie lange man in die errechnete Richtung wandert hängt auch davon ab, wie man z.B. die Schrittweite ermittelt, dazu später mehr.

In den folgenden Absätzen werde ich versuchen zwei verschiedene Arten des Gradientenverfahrens zu zeigen und zu erklären.

2 Schrittweitenbestimmung mittels Gleichungssystem

In diesem Abschnitt möchte ich zeigen, wie man relative einfach (zu nicht zu komplizierten Funktionen) mittels dem Gradientenverfahren ein Minimum findet.

2.1 Der Ablauf

1. Es wird ein beliebiger Startpunkt $P = (x, y)$ ausgewählt.
2. Es wird die Ableitung der Funktion $f(x, y)$, also $\nabla f(x)$ gebildet. Sollte

dieser Gradient bereits den Wert null haben sind wir bereits fertig und haben ein lokales Minimum.

3. Der Punkt P wird in den Gradienten eingesetzt und der negative Wert berechnet, also $-\nabla f(x_P, y_P)$.
4. Um einen neuen Punkt $P^{[n+1]}$ zu finden stellen wir nun folgende Gleichung auf: $P^{[n+1]} = P^{[n]} + \alpha_{[n]} * s^{[n]}$. Nun haben wir die Richtung in die optimiert werden muss gefunden ($s^{[n]} (= -\nabla f(x_P, y_P))$), jetzt muss noch entschieden werden wie weit wir in diese Richtung "wandern" ($\alpha^{[n]}$, liegt zwischen 0 und 1) (Hinweis: n ist die aktuelle Iteration).

In diesem Kapitel werden wir die genau richtige Schrittweite auf Anhieb berechnen, dafür werden schauen wir uns zuerst die obere Gleichung $P^{[n+1]} = P^{[n]} + \alpha_{[n]} * s^{[n]}$ an. Aus dieser kann man herauslesen: $x = P_x^{[n]} + \alpha_{[n]} * s_x^{[n]}$ und $y = P_y^{[n]} + \alpha_{[n]} * s_y^{[n]}$. Dieses x und y wird nun in f eingesetzt und die daraus resultierende Funktion, welche nur noch von α abhängig ist abermals abgeleitet.

Diese abgeleitete Funktion wird nun gleich null gesetzt und so das passende α errechnet. Diese Art ein *alpha* zu errechnen kann zwar bei einfachen Funktionen recht effizient von statten gehen, bei sehr komplexen und hochdimensionalen Funktionen aber auch enorm kompliziert werden, im gegensatz zu der in so einem Fall etwas simpleren Schrittweitenbestimmung mittels Armijo-Bedingung.

5. Wenn man einen neuen Punkt gefunden hat wird bei Punkt 3 fortgesetzt. Da es oft einen großen Rechenaufwand erfordert das genaue Minimum zu finden wird das Verfahren meist beendet wenn der Gradient einen akzeptablen Wert erreicht hat (So kann man als Abbruchbedingung z.B. die stärke des Anstiegs in einem Punkt (also den Betrag des Gradienten) nehmen).

2.2 Beispiel

Zu minimierende Funktion: $f(x, y) = 3x^3 + y^2 - 3xy - 3x$, als Abbruchbedingung wählen wir $|\nabla f(x, y)| \leq 0.6$

1. Startpunkt $P = (1, 1)$
2. Ableitung: $\nabla f(x, y) = \begin{pmatrix} 6x - 3y - 3 \\ 2y - 3x \end{pmatrix}$ (\rightarrow Der Gradient)
3. $-\nabla f(1, 1) = \begin{pmatrix} 6 - 3 - 3 \\ 2 - 3 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$ (\rightarrow Richtung des stärksten Abstiegs)

4. Nun haben wir folgende Gleichung: $P^{[1]} = P^{[0]} + \alpha * -\nabla f(P_x^{[0]}, P_y^{[0]})$
 bzw. $P^{[1]} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \alpha * \begin{pmatrix} 0 \\ -1 \end{pmatrix}$. Das heißt unser $x = 1$ und unser
 $y = 1 - \alpha$. Diese Werte werden nun in f eingesetzt, sodass: $f(\alpha) = 3 + (1 - \alpha)^2 - 3(1 - \alpha) - 3$ bzw. $f(\alpha) = (1 - \alpha)^2 - 3(1 - \alpha)$ und abgeleitet,
 also: $f'(\alpha) = 3 - 2 * (1 - \alpha)$. Wenn wir die abgeleitete Funktion gleich
 null setzen und die daraus resultierende Gleichung lösen bekommen
 wir das optimale $\alpha = -0.5$. Nun haben wir ein passendes α gefunden
 und können zum nächsten Punkt wandern.

$$5. P^{[1]} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 0.5 * \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1.5 \end{pmatrix}$$

$$6. -\nabla f(1, 1.5) = \begin{pmatrix} 1.5 \\ 0 \end{pmatrix}$$

$$7. P^{[2]} = \begin{pmatrix} 1 \\ 1.5 \end{pmatrix} + \alpha * \begin{pmatrix} 1.5 \\ 0 \end{pmatrix}$$

$\rightarrow x = 1 + 1.5\alpha, y = 1.5$
 $\rightarrow f(\alpha) = 3(1 + 1.5\alpha)^2 + 1.5^2 - 4.5(1 + 1.5\alpha) - 3(1 + 1.5\alpha)$
 $\rightarrow f'(\alpha) = \frac{54\alpha - 9}{4}$
 $\rightarrow \alpha = \frac{1}{6}$

$$8. P^{[2]} = \begin{pmatrix} 1 \\ 1.5 \end{pmatrix} + \frac{1}{6} * \begin{pmatrix} 1.5 \\ 0 \end{pmatrix} = \begin{pmatrix} 1.25 \\ 1.5 \end{pmatrix}$$

$$9. -\nabla f(1.25, 1.5) = \begin{pmatrix} 0 \\ 0.75 \end{pmatrix}$$

$$10. P^{[3]} = \begin{pmatrix} 1.25 \\ 1.5 \end{pmatrix} + \alpha * \begin{pmatrix} 0 \\ 0.75 \end{pmatrix}$$

$\rightarrow x = 1.25, y = 1.5 + 0.75\alpha$
 $\rightarrow f(\alpha) = 3 * 1.25^2 + (1.5 + 0.75\alpha)^2 - 3.75(1.5 + 0.75\alpha) - 3.75$
 $\rightarrow f'(\alpha) = \frac{18\alpha - 9}{16}$
 $\rightarrow \alpha = 0.5$

$$11. P^{[3]} = \begin{pmatrix} 1.25 \\ 1.5 \end{pmatrix} + 0.5 * \begin{pmatrix} 0 \\ 0.75 \end{pmatrix} = \begin{pmatrix} 1.25 \\ 1.875 \end{pmatrix}$$

$$12. -\nabla f(1.25, 1.875) = \begin{pmatrix} 1.125 \\ 0 \end{pmatrix}$$

$$13. P^{[4]} = \begin{pmatrix} 1.25 \\ 1.875 \end{pmatrix} + \alpha * \begin{pmatrix} 1.125 \\ 0 \end{pmatrix}$$

$\rightarrow x = 1.25 + 1.125\alpha, y = 1.875$

$$\begin{aligned}
&\rightarrow f(\alpha) = 3 * (1.25 + 1.125\alpha)^2 + 1.875^2 - 3 * 1.875 * (1.25 + 1.125\alpha) - \\
&3 * (1.25 + 1.125\alpha) \\
&\rightarrow f'(\alpha) = \frac{18\alpha-9}{16} \\
&\rightarrow \alpha = 0.5
\end{aligned}$$

$$14. P^{[4]} = \begin{pmatrix} 1.25 \\ 1.875 \end{pmatrix} + 0.5 * \begin{pmatrix} 1.125 \\ 0 \end{pmatrix} = \begin{pmatrix} 1.438 \\ 1.875 \end{pmatrix}$$

$$15. -\nabla f(1.438, 1.875) = \begin{pmatrix} 0 \\ 0.563 \end{pmatrix}$$

16. Da $|\nabla f(1.438, 1.875)| = \sqrt{0^2 + 0.563^2} = 0.563$ und $0.563 < 0.6$ haben wir einen Punkt gefunden, welcher nahe genug am Optimum liegt.

3 Schrittweitenbestimmung mittels der Armijo-Bedingung

3.1 Der Ablauf

1. Es wird ein beliebiger Startpunkt $P = (x, y)$ ausgewählt.
2. Es wird die Ableitung der Funktion $f(x, y)$, also $\nabla f(x)$ gebildet. Sollte dieser Gradient bereits den Wert null haben sind wir bereits fertig und haben ein lokales Minimum.
3. Der Punkt P wird in den Gradienten eingesetzt und der negative Wert berechnet, also $-\nabla f(x_P, y_P)$.
4. Um einen neuen Punkt $P^{[n+1]}$ zu finden stellen wir nun folgende Gleichung auf: $P^{[n+1]} = P^{[n]} + \alpha_{[n]} * s^{[n]}$. Nun haben wir die Richtung in die optimiert werden muss gefunden ($s^{[n]} (= -\nabla f(x_P, y_P))$), jetzt muss noch entschieden werden wie weit wir in diese Richtung "wandern" ($\alpha^{[n]}$, liegt zwischen 0 und 1) (Hinweis: n ist die aktuelle Iteration).

Das kann man einerseits mit Hilfe eines fixen Werts machen, andererseits auch mit einer Folge (wie z.B. $\frac{1}{\sqrt{n}}$) oder einer jedes mal neu errechneten Schrittweite. Zu beachten ist, dass die ausgewählte Folge auf jeden Fall divergent sein muss, da man sonst bis zu einer maximalen Entfernung "wandern" kann und die Wahrscheinlichkeit groß ist das Optimum nie zu erreichen. Sollte man jedes mal einen fixen Wert verliert man zwar den Aufwand für die Berechnung der Schrittweite, jedoch ist die Wahrscheinlichkeit hoch, dass man viele Iterationen benötigt oder in einer endlosen Schleife festhängt.

Es gibt verschieden Arten wie man die Schrittweite berechnen kann, im

Folgenden werde ich die Schrittweitenbestimmung nach Armijo erklären. Dieses Verfahren ist eines der sogenannten "line-search" Verfahren, das von mir gezeigt ist jedoch kein exaktes line-search Verfahren sondern lediglich eine einfache Heuristik.

Die Armijo-Bedingung: $\varphi(\alpha) \leq \varphi(0) + c * \alpha^{[n]} * \varphi(0)'$.

Wobei c eine Konstante zwischen null und eins ist und dazu dient das die Bedingung nicht zu restriktiv ist, $\varphi(\alpha) = f(P^{[n]} + \alpha^{[n]} s^{[n]})$ und $\varphi(0)' = \nabla f(P^n) * s^{[n]}$

5. Für den Anfang wird $\alpha = 1$ gewählt und errechnet ob die Armijo-Bedingung erfüllt ist - sollte sie erfüllt sein haben wir ein passende α gefunden, wenn nicht wird α mit einem β , welches ebenfalls zwischen null und eins liegt, multipliziert und wieder geprüft ob die Ungleichung erfüllt ist. (Hinweis: In der Praxis liefern die Werte $c = 0.01$ und $\beta = 0.9$ häufig gute Ergebnisse. (QUELLE: <https://www.unibw.de/lrt1/gerdts/lehre/lpnlp/lpnlp.pdf>)
6. Wenn man einen neuen Punkt gefunden hat wird bei Punkt 3 fortgesetzt.

3.2 Beispiel

Zu minimierende Funktion: $f(x, y) = y^4 + 2x^2 - 3xy + 1$, als Abbruchbedingung wählen wir $|\nabla f(x, y)| \leq 0.3$

1. Startpunkt $P = (1, 1)$
2. Ableitung: $\nabla f(x, y) = \begin{pmatrix} 4x - 3y \\ 4y^3 - 3x \end{pmatrix}$ (\rightarrow Der Gradient)
3. $-\nabla f(1, 1) = \begin{pmatrix} -4 + 3 \\ -4 + 3 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$ (\rightarrow Richtung des stärksten Abstiegs)
4. Die Armijo-Bedingung (mit $\alpha = 1$, $c = 0.1$ und $\beta = 0.9$):

$$f\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} + 1 * \begin{pmatrix} -1 \\ -1 \end{pmatrix}\right) \leq f\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) + 0.1 * 1 * \begin{pmatrix} 1 \\ 1 \end{pmatrix} * \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

$$= 1 \leq 0.8$$

$$\rightarrow \alpha^{[2]} = \alpha^{[1]} * \beta = 0.9$$

$$f\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} + 0.9 * \begin{pmatrix} -1 \\ -1 \end{pmatrix}\right) \leq f\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) + 0.1 * 0.9 * \begin{pmatrix} 1 \\ 1 \end{pmatrix} * \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

$$= 0.99 \leq 0.82$$

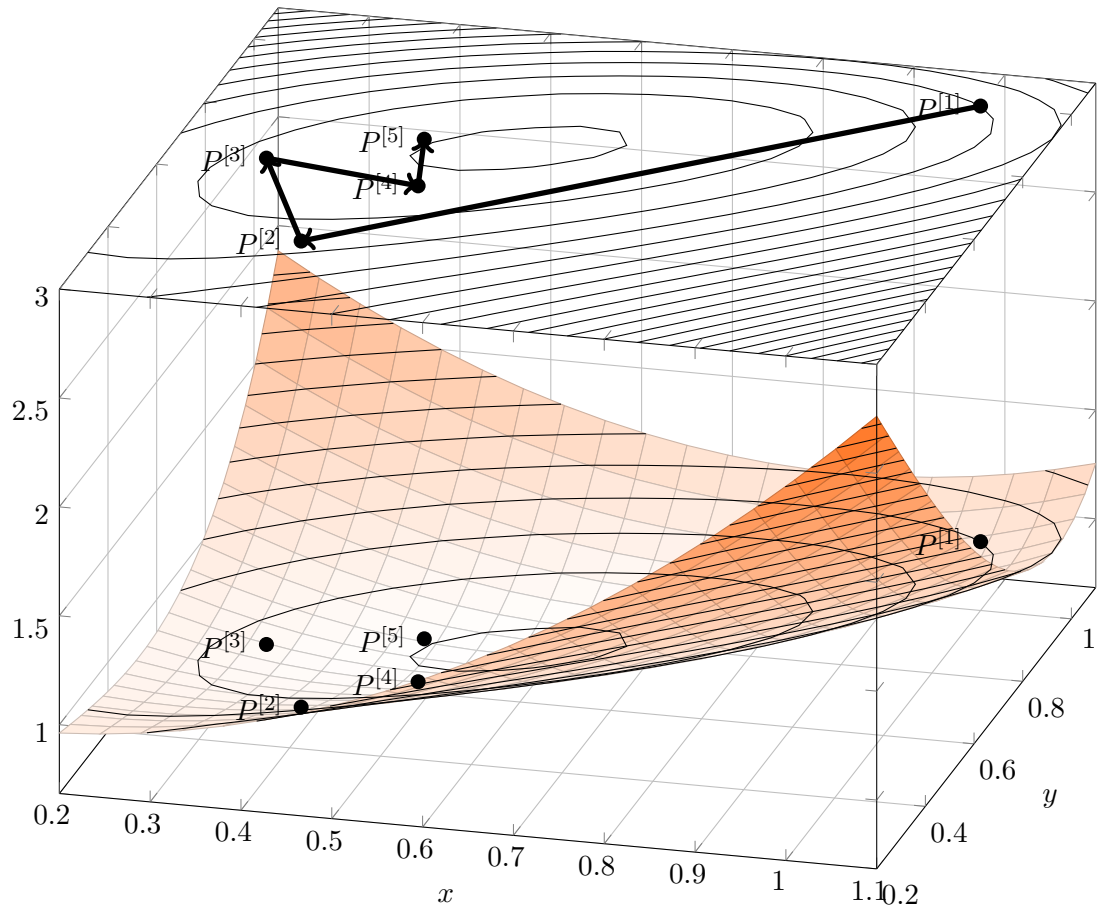


Abbildung 2: Hier zu sehen ist die Zweidimensionale Funktion, die bei der Angabe von Beispiel 2 gegeben ist. Zusätzlich sind hier die Höhenschichtlinien eingezeichnet. Wie man sieht nähert man sich mit jeder Iteration an das Optimum der Funktion an.

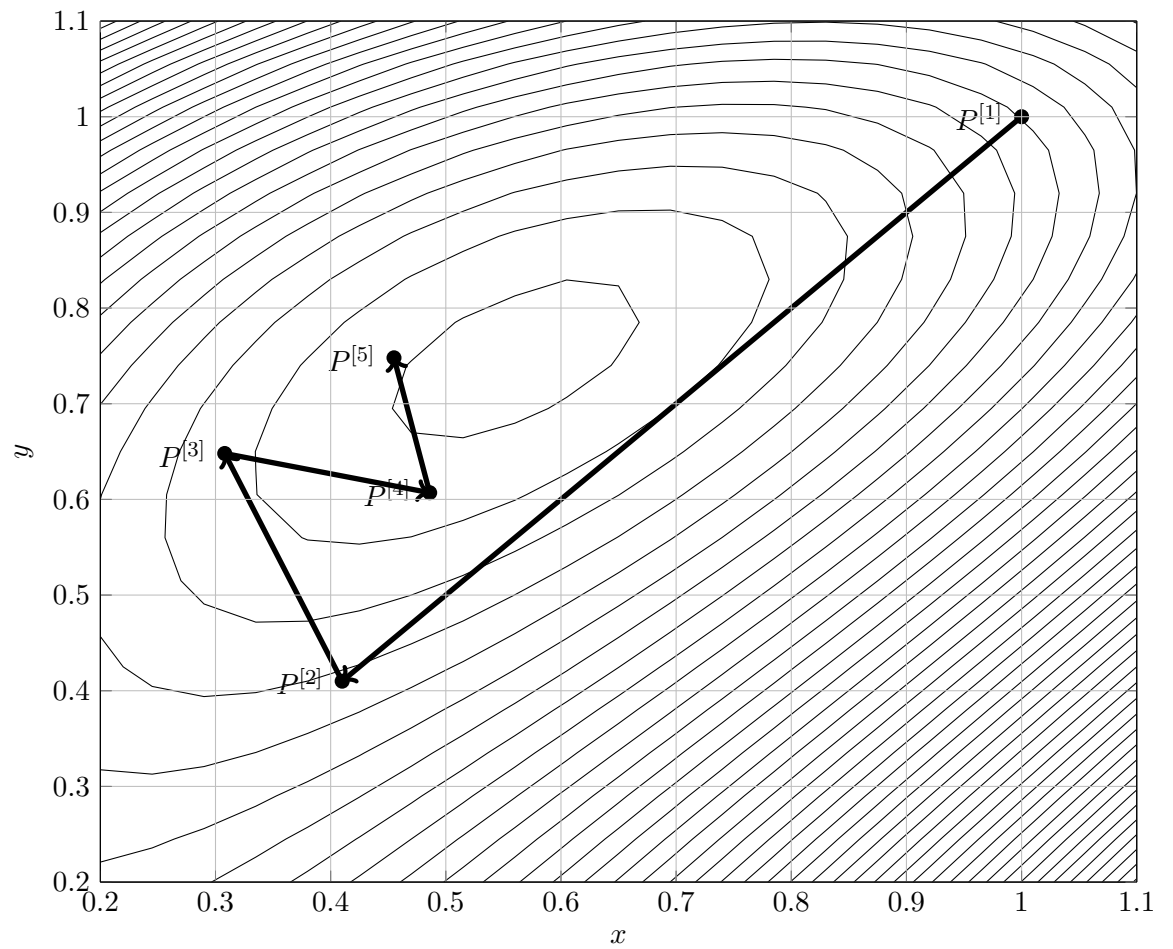


Abbildung 3: Hier zu sehen sind die Höhenlinien der Funktion bei Beispiel 1. Wie man sieht nähert man sich mit jeder Iteration an das Optimum der Funktion an.

$$\rightarrow \alpha^{[3]} = \alpha^{[2]} * \beta = 0.81$$

$$\begin{aligned} f\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} + 0.81 * \begin{pmatrix} -1 \\ -1 \end{pmatrix}\right) &\leq f\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) + 0.1 * 0.81 * \begin{pmatrix} 1 \\ 1 \end{pmatrix} * \begin{pmatrix} -1 \\ -1 \end{pmatrix} \\ &= 0.965 \leq 0.838 \\ \rightarrow \alpha^{[4]} &= \alpha^{[3]} * \beta = 0.729 \end{aligned}$$

$$\begin{aligned} f\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} + 0.729 * \begin{pmatrix} -1 \\ -1 \end{pmatrix}\right) &\leq f\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) + 0.1 * 0.729 * \begin{pmatrix} 1 \\ 1 \end{pmatrix} * \begin{pmatrix} -1 \\ -1 \end{pmatrix} \\ &= 0.932 \leq 0.854 \\ \rightarrow \alpha^{[5]} &= \alpha^{[4]} * \beta = 0.656 \end{aligned}$$

$$\begin{aligned} f\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} + 0.656 * \begin{pmatrix} -1 \\ -1 \end{pmatrix}\right) &\leq f\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) + 0.1 * 0.656 * \begin{pmatrix} 1 \\ 1 \end{pmatrix} * \begin{pmatrix} -1 \\ -1 \end{pmatrix} \\ &= 0.896 \leq 0.869 \\ \rightarrow \alpha^{[6]} &= \alpha^{[5]} * \beta = 0,59 \end{aligned}$$

$$\begin{aligned} f\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} + 0,59 * \begin{pmatrix} -1 \\ -1 \end{pmatrix}\right) &\leq f\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) + 0.1 * 0,59 * \begin{pmatrix} 1 \\ 1 \end{pmatrix} * \begin{pmatrix} -1 \\ -1 \end{pmatrix} \\ &= 0.86 \leq 0.882 \end{aligned}$$

Nun haben wir ein passendes α gefunden und können zum nächsten Punkt wandern.

$$5. P^{[2]} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 0,59 * \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0.41 \\ 0.41 \end{pmatrix}$$

$$\begin{aligned} 6. -\nabla f(0.41, 0.41) &= \begin{pmatrix} -0.41 \\ 0.95 \end{pmatrix}, \\ |\nabla f(0.41, 0.41)| &= \sqrt{0.41^2 + 0.95^2} = 1.071 \end{aligned}$$

7. Die Armijo-Bedingung (mit $\alpha = 1$, $c = 0.1$ und $\beta = 0.5$):

$$\begin{aligned} f\left(\begin{pmatrix} 0.41 \\ 0.41 \end{pmatrix} + 1 * \begin{pmatrix} -0.41 \\ 0.95 \end{pmatrix}\right) &\leq f\left(\begin{pmatrix} 0.41 \\ 0.41 \end{pmatrix}\right) + 0.1 * 1 * \begin{pmatrix} -0.41 \\ 0.95 \end{pmatrix} * \\ &\begin{pmatrix} 0.41 \\ -0.95 \end{pmatrix} \\ &= 4.421 \leq 0.753 \rightarrow \alpha^{[2]} = \alpha^{[1]} * \beta = 0.5 \\ &\rightarrow = 1.153 \leq 0.807 \rightarrow \alpha^{[3]} = \alpha^{[2]} * \beta = 0.25 \\ &\rightarrow = 0.768 \leq 0.833 \end{aligned}$$

$$8. P^{[3]} = \begin{pmatrix} 0.41 \\ 0.41 \end{pmatrix} + 0.25 * \begin{pmatrix} -0.41 \\ 0.95 \end{pmatrix} = \begin{pmatrix} 0.308 \\ 0.648 \end{pmatrix}$$

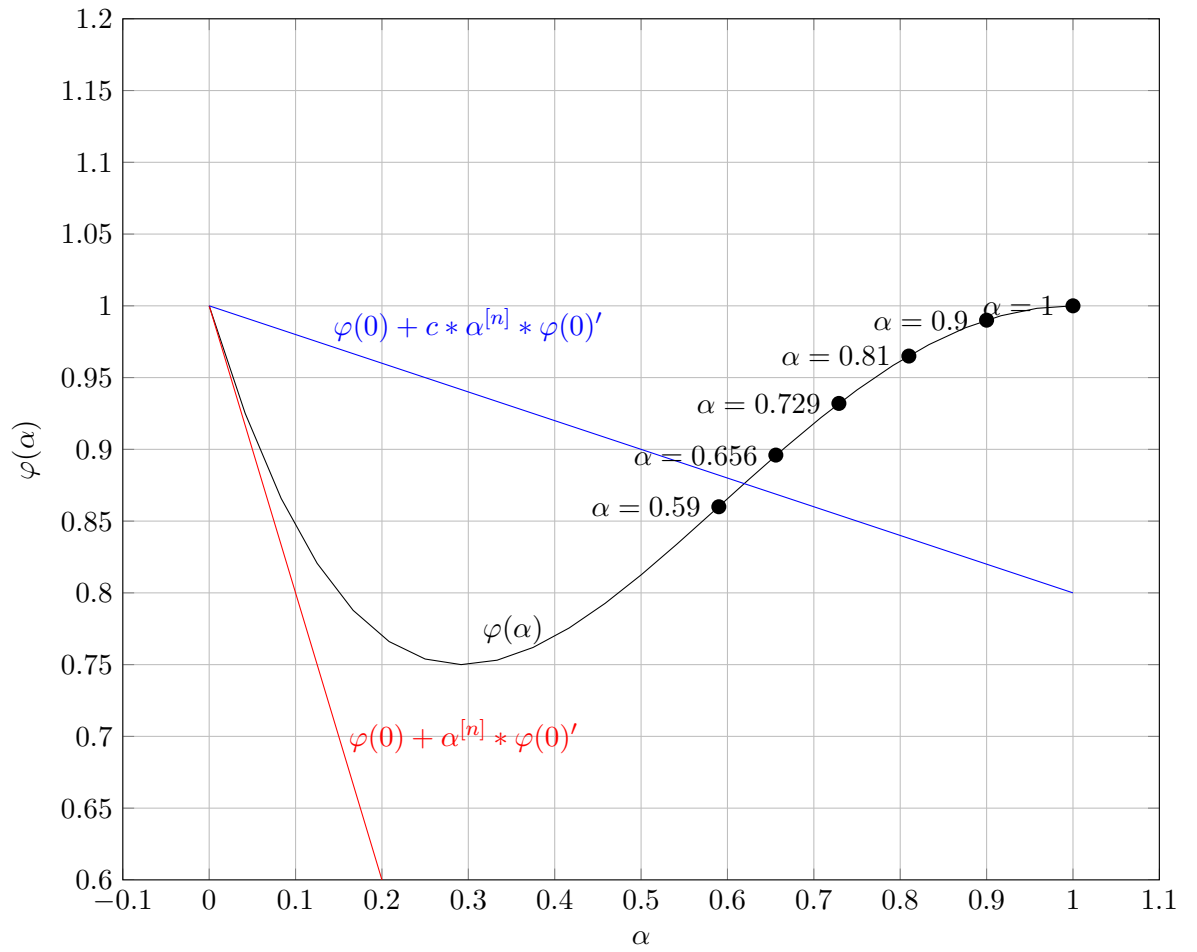


Abbildung 4: Hier sieht man die Funktion $\varphi(\alpha) = (1 - \alpha)^4 + 2 * (1 - \alpha)^2 - 3 * (1 - \alpha) * (1 - \alpha) + 1$, sie ist sozusagen ein Durchschnitt durch die bei Beispiel 1 gegebene Funktion. Diese Funktion wird erzeugt, wenn man vom Punkt $P^{[1]} = (1, 1)$ in Richtung des Gradienten wandert. Die blaue Gerade zeigt die Armijo-Bedingung bei $c = 0.1$, die rote bei $c = 1$ bzw. ohne c .

$$9. -\nabla f(0.308, 0.648) = \begin{pmatrix} 0.712 \\ -0.164 \end{pmatrix},$$

$$|\nabla f(0.308, 0.648)| = \sqrt{0.712^2 + 0.164^2} = 0.534$$

10. Die Armijo-Bedingung (mit $\alpha = 1$, $c = 0.1$ und $\beta = 0.5$):

$$f\left(\begin{pmatrix} 0.308 \\ 0.648 \end{pmatrix} + 1 * \begin{pmatrix} 0.712 \\ -0.164 \end{pmatrix}\right) \leq f\left(\begin{pmatrix} 0.308 \\ 0.648 \end{pmatrix}\right) + 0.1 * 1 * \begin{pmatrix} 0.712 \\ -0.164 \end{pmatrix} *$$

$$\begin{pmatrix} -0.712 \\ 0.164 \end{pmatrix}$$

$$= 1.655 \leq 0.714 \rightarrow \alpha^{[2]} = \alpha^{[1]} * \beta = 0.5$$

$$\rightarrow = 0.857 \leq 0.741 \rightarrow \alpha^{[3]} = \alpha^{[2]} * \beta = 0.25$$

$$\rightarrow = 0.723 \leq 0.754$$

$$11. P^{[4]} = \begin{pmatrix} 0.308 \\ 0.648 \end{pmatrix} + 0.25 * \begin{pmatrix} 0.712 \\ -0.164 \end{pmatrix} = \begin{pmatrix} 0.486 \\ 0.607 \end{pmatrix}$$

$$12. -\nabla f(0.486, 0.607) = \begin{pmatrix} -0.123 \\ 0.563 \end{pmatrix},$$

$$|\nabla f(0.486, 0.607)| = \sqrt{0.123^2 + 0.563^2} = 0.332$$

13. Die Armijo-Bedingung (mit $\alpha = 1$, $c = 0.1$ und $\beta = 0.5$):

$$f\left(\begin{pmatrix} 0.486 \\ 0.607 \end{pmatrix} + 1 * \begin{pmatrix} -0.123 \\ 0.563 \end{pmatrix}\right) \leq f\left(\begin{pmatrix} 0.486 \\ 0.607 \end{pmatrix}\right) + 0.1 * 1 * \begin{pmatrix} -0.123 \\ 0.563 \end{pmatrix} *$$

$$\begin{pmatrix} 0.123 \\ -0.563 \end{pmatrix}$$

$$= 1.863 \leq 0.689 \rightarrow \alpha^{[2]} = \alpha^{[1]} * \beta = 0.5$$

$$\rightarrow = 0.852 \leq 0.707 \rightarrow \alpha^{[3]} = \alpha^{[2]} * \beta = 0.25$$

$$\rightarrow = 0.706 \leq 0.715$$

$$14. P^{[5]} = \begin{pmatrix} 0.486 \\ 0.607 \end{pmatrix} + 0.25 * \begin{pmatrix} -0.123 \\ 0.563 \end{pmatrix} = \begin{pmatrix} 0.455 \\ 0.748 \end{pmatrix}$$

$$15. -\nabla f(0.455, 0.748) = \begin{pmatrix} 0.424 \\ -0.309 \end{pmatrix},$$

$$|\nabla f(0.455, 0.748)| = \sqrt{0.424^2 + 0.309^2} = 0.275$$

Die Abbruchbedingung ist erfüllt und somit ist ein Punkt, welcher nahe genug am Optimum liegt gefunden.