ii) Nun betrachten wir die folgende Funktion: $g(x,y) = \tan(xy)$ auf dem offenen Gebiet $G = (0,\pi/2)\times(0,1)$. Zuerst suchen wir kritische Punkte:

$$\nabla g(x,y) = \begin{pmatrix} (1+\tan^2(xy))y \\ (1+\tan^2(xy))x \end{pmatrix} \stackrel{!}{=} 0$$

hat nur die Lösung $x = y = 0$. Dieser Punkt gehört aber nicht zum betrachteten Gebiet. Wir haben also in unserem Fall keine kritischen Punkte.
Als Nächstes müssen wir uns überlegen, was passiert, wenn wir uns dem Rand des gegebenen Gebiets nähern. Dabei gilt das Folgende: Alle möglichen Grenzwerte, wenn wir uns dem Rand von G nähern, sind null, ausser dem folgenden:

$$\lim_{(x,y)\to(\pi/2,1)} g(x,y) = \infty.$$

Wiederum haben wir eine Situation, in welcher weder das globale Minimum noch das globale Maximum angenommen werden.

iii) Nun betrachten wir die Funktion $h(x,y) = \cos(x^2+y^2)$ auf der offenen Einheitskreisscheibe $D = \{(r,\varphi) \mid 0 \leq r < 1\}$. Hier kann man sich leicht davon überzeugen, dass das globale Maximum im Ursprung angenommen wird, das globale Mininum hingegen nicht angenommen wird, da die Kreislinie $r = 1$ nicht zur Einheitskreisscheibe dazugehört.

Wir fassen das Vorgehen also wie folgt zusammen:

Vorgehen zur Bestimmung optimaler Werte auf einem offenen Gebiet:

- Bestimmung und Diskussion aller kritischen Punkte
- Bestimmung der Punkte, an denen der Gradient nicht definiert ist
- Liste der Funktionswerte an den Punkten, die in den beiden vorangehenden Schritten identifiziert wurden
- Was passiert, wenn wir uns dem Rand des betrachteten Gebiets, respektive allfälligen Definitionslücken nähern? Berechnung allfälliger Grenzwerte
- Evaluation der zusammengetragenen Informationen. Insbesondere auch die Frage: Welche Werte werden tatsächlich angenommen?

1.5.1 Methode der Lagrange-Multiplikatoren

Im Folgenden betrachten wir die spezielle Situation, dass wir eine Extremalstelle einer Funktion $f(x,y)$ suchen, wobei aber die Menge aller möglichen Punkte, die wir in Betracht ziehen dürfen, **nicht die ganze xy-Ebene ist**, sondern lediglich die Menge derjenigen Punkte ist, welche eine **zusätzliche Nebenbedingung** erfüllen.

Probleme dieser Form tauchen in vielen verschiedenen Kontexten auf, z.B.

- Optimierung einer Wirkstoffkombination zweier Medikamente, wobei die Gesamtmedikamentenmenge begrenzt ist.
- Beste Termineinteilung medizinischer Kontrollen bei fixem Arbeitspensum. [1]

[1] vgl. auch Daniel Kent, Ross Shachter, et al.; Efficient Scheduling of Cystoscopies in Monitoring for Recurrent Bladder Cancer in Medical Decision Making (Philadelphia: Hanley and Belfus, 1989)

- Optimierung einer Produktion zweier Produkte unter begrenzter Investition.
- Berücksichtigung physikalischer Rahmenbedingungen

Wir wollen uns diese Situation an einem ganz konkreten Beispiel anschauen.

Beispiel 1.23. Gegeben ist die Funktion

$$f(x,y) = -2x + y + 1 \tag{1.4}$$

Gesucht sind Punkte, an welchen f einen optimalen Wert annimmt und welche **zusätzlich** die **Nebenbedingung**

$$y - x^2 = 0 \tag{1.5}$$

erfüllen.

Erste, einfache Lösungsstrategie:
Die Nebenbedingung kann nach y aufgelöst werden

$$y = x^2$$

Dies kann nun in der Formel der zu optimierenden Funktion eingesetzt werden

$$f(x,y) = -2x + x^2 + 1.$$

Dies bedeutet aber nichts anderes, als dass wir nun eine Funktion in **einer** Variablen optimieren müssen, nämlich

$$h(x) = -2x + x^2 + 1.$$

Wir suchen also nach kritischen Punkten von h. Dazu muss gelten $h'(x) = 0$, also

$$h'(x) = -2 + 2x = 0.$$

Für die Lösung dieser Gleichung erhalten wir $x = 1$ und damit $y = 1$. Damit haben wir einen **Kandidaten** für eine Extremalstelle unter der zusätzlichen Nebenbedingung (1.5).
An diesem Punkt $(1,1)$ gilt $f(1,1) = 0$.
Durch Vergleich mit anderen Punkten in einer Umgebung von $(1,1)$, welche die Nebenbedingung ebenfalls erfüllen, also von der Form (x, x^2) sind, sehen wir, dass an der Stelle $(1,1)$ das Minimum von f unter der Nebenbedingung $y - x^2 = 0$ angenommen wird.

Wie kann das Gesehene etwas allgemeiner formuliert werden?
Unser Vorgehen, dass wir zuerst die Nebenbedingung nach y aufgelöst haben und dieses Resultat dann in die Formel der zu optimierenden Funktion eingesetzt haben, bedeutet geometrisch, dass wir uns zuerst auf die Lösungsmenge der Gleichung der Nebenbedingung, also auf die Parabel $y = x^2$, begeben und dann in einem zweiten Schritt nach einem Optimum der Funktion f auf dieser Kurve suchen.
Mit anderen Worten: Durch die Nebenbedingung schränken wir die Menge aller betrachteten Punkte, wo eine Extremalstelle allenfalls auftreten kann, ein; wir betrachten nämlich nicht mehr alle Punkte in der xy-Ebene, sondern nur noch Punkte, welche die Gleichung $y - x^2 = 0$ erfüllen.

Als Nächstes giessen wir diese Idee in Formeln:
Die Punkte auf dieser Parabel können wir wie folgt parametrisieren:

$$\gamma : t \mapsto \begin{pmatrix} t \\ t^2 \end{pmatrix} = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}$$

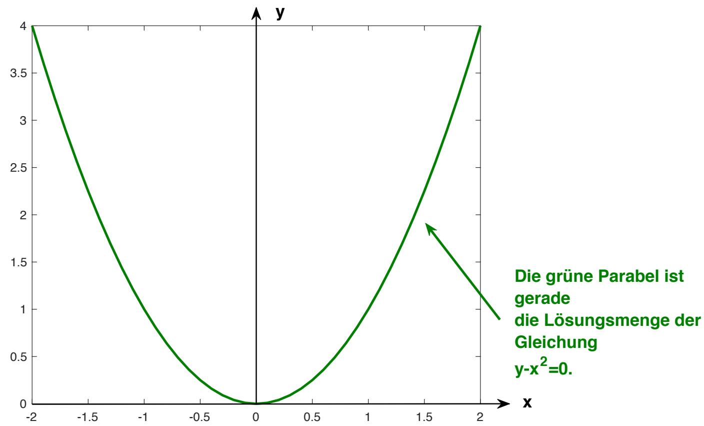

Abbildung 1.27: Lösungsmenge der Nebenbedingung $y - x^2 = 0$

In einem zweiten Schritt überlegen wir uns, wie unsere Funktion f, die wir optimieren wollen, entlang dieser Parametrisierung aussieht. Dies ist nichts anderes, als dass wir die **zusammengesetzte Funktion**

$$f(\gamma(t)) = f(x(t), y(t))$$

anschauen.
Diese zusammengesetzte Funktion $f(\gamma(t))$ muss nun optimiert werden. Dazu suchen wir die Lösungen der Gleichung (kritische Punkte)

$$\frac{d}{dt} f(x(t), y(t)) = 0$$

Die **Kettenregel** - formuliert mit Hilfe des Gradienten - ergibt

$$\frac{d}{dt} f(x(t), y(t)) = \nabla f \cdot \begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2t \end{pmatrix} = -2 + 2t = 0 \tag{1.6}$$

Der einzige kritische Punkt ist also $t = 1$, d.h $(1, 1)$ in der xy-Ebene.

Geometrisch bedeutet die Bedingung (1.6), dass in einem optimalen Punkt, welcher insbesondere die Nebenbedingung erfüllt, die beiden Vektoren

$$\nabla f \quad \text{und} \quad \begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} \quad \text{senkrecht aufeinander stehen.}$$

Diese Situation mitsamt den soeben zusammengetragenen Beobachtungen ist unten geometrisch illustriert.

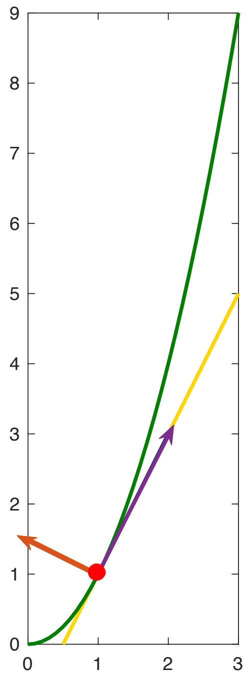

Abbildung 1.28: Geometrische Situation

Zusammenfassung:

- Der Gradient der zu optimierenden Funktion f steht senkrecht auf der Tangente der Nebenbedingung.
- Da wir wissen, dass der Gradient von f stets senkrecht auf der Höhenlinie durch den betrachteten Punkt steht, gilt nun, dass die Tangente der Nebenbedingung im kritischen Punkt - unter der gegebenen Nebenbedingung - mit einer Höhenlinie der zu optimierenden Funktion zusammenfallen muss.
- Die gegebene Nebenbedingung $y - x^2 = 0$ kann auch als Niveaulinie der Funktion $g(x, y) = y - x^2$ zum Niveau null aufgefasst werden. Da der Gradient von g ebenfalls auf der Niveaulinie senkrecht steht, muss nun also gelten
$$\nabla g \quad \text{und} \quad \nabla f \quad \text{sind parallel}$$
also
$$\nabla f = \lambda \nabla g$$
für eine reelle Zahl λ.
Diese letzte Beobachtung erklärt die **Methode der Lagrange-Multiplikatoren**.

Bemerkung:
Die Situation kann **geometrisch** wie folgt beschrieben werden: Im Definitionsbereich der zu optimieren Funktion $f(x, y)$ (in unserem Fall also die xy-Ebene, in der untenstehenden Graphik grün dargestellt) schränken wir die Punkte, welche wir noch betrachten, ein, nämlich auf diejenigen Punkte, welche zudem die Nebenbedingung $g(x, y) = 0$ erfüllen. Übrigt bleit also eine Kurve im Definitionsbereich (in der Abbildung unten blau eingezeichnet).
Im drei-dimensionalen Raum entspricht dies der Tatsache, dass nicht mehr der gesamte Graph der zu optimieren Funktion $f(x, y)$ (in der Abbildung gelb dargestellt) betrachtet wird, sondern nur eine Kurve (pink in der Abbildung), welche auf diesem Graphen liegt. In unserem Fall können diese Punkte beschrieben werden durch $(x, x^2, -2x + x^2 + 1)$.
Die bedingte Optimierung kann also geometrisch so beschrieben werden, dass wir danach fragen, an welchem Punkt auf der pinken Kurve der kleinste z-Wert angenommen und wo der grösste z-Wert angenommen wird.

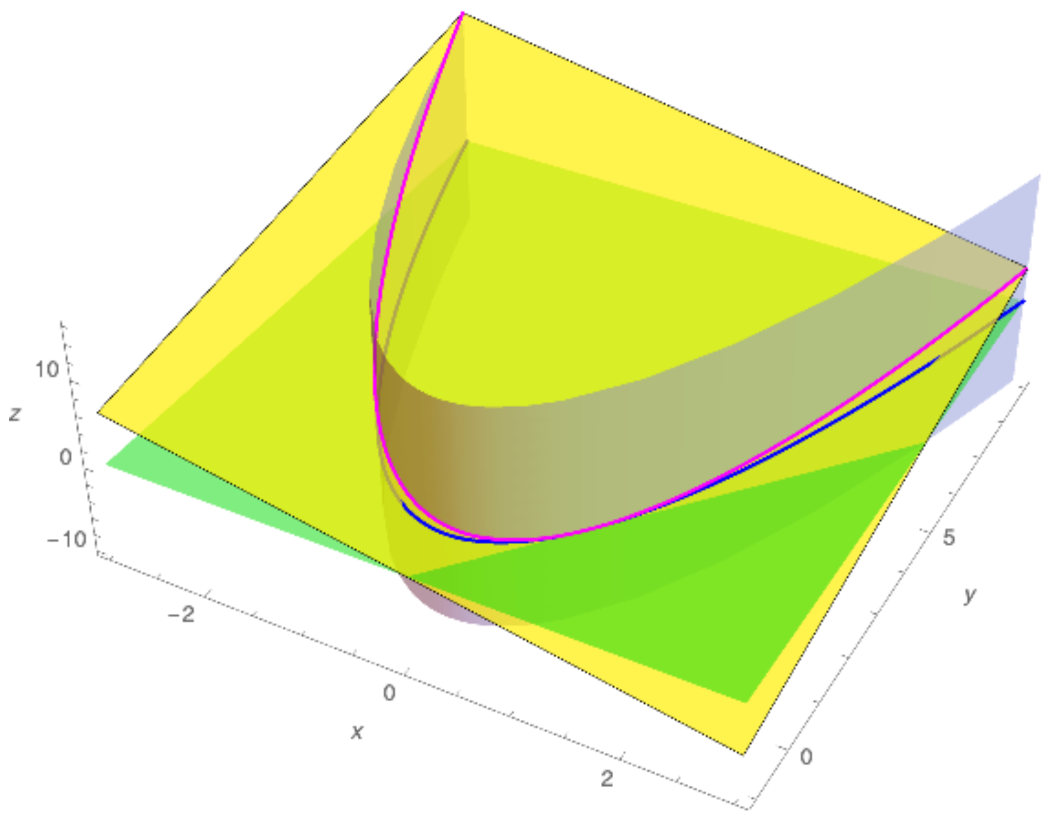

Abbildung 1.29: Illustration zur bedingten Optimierung

Was wir in diesem Beispiel gesehen haben, gilt allgemein und wird als Methode der Lagrange-Multiplikatoren bezeichnet. Wir haben also die folgende Vorgehensweise: