3.7 Constrained critical points and Lagrange multipliers

**Theorem 3.7.1.** Let $M \subset \mathbb{R}^n$ be a $k$-dimensional manifold, let $U \subset \mathbb{R}^n$ be an open set, and let $f : U \to \mathbb{R}$ be a $C^1$ function. If $c \in M \cap U$ is a local extremum of $f$ restricted to $M$, then

$$T_c M \subset \ker [Df_c].$$

That is, the gradient $\nabla f_c$ is perpendicular to $T_c M$ (because $Df_c(v) = \nabla f_c \cdot v$ tells us that if $v \in T_c M$, then $T_c M \subset \ker [Df_c]$ implies $\nabla f_c \cdot v = Df_c(v) = 0$).

**Proof.** Near $c$, $M$ is the graph of a $C^1$ function $g : V \to \mathbb{R}^k$

$$g(x) = f(x) = (x, g(x)) : x \in V = M \cap \text{open set}.$$  

Since $c \in M \cap U$ is a local extremum of $f$ restricted to $M$, $\gamma^{-1}(c)$ is a local extremum of the function $f \circ \gamma : V \to \mathbb{R}$.

Thus

$$0 = D(f \circ \gamma)_{\gamma^{-1}(c)} = Df_c \circ D\gamma_{\gamma^{-1}(c)}.$$

Hence

$$T_c M = \text{img}(D\gamma_{\gamma^{-1}(c)}) \subset \ker(Df_c).$$

Here’s another proof of the theorem: Let $v \in T_c M$. Then there exists a curve $\alpha : (-\varepsilon, \varepsilon) \to M$ such that $\alpha(0) = c$ and $\alpha'(0) = v$. Since $c$ is a local extremum of $f$ restricted to $M$ and since $\alpha(0) = c$, the function $f \circ \alpha : (-\varepsilon, \varepsilon) \to \mathbb{R}$ has a local extremum at 0. Hence

$$0 = (f \circ \alpha)'(0) = Df_{\alpha(0)}(\alpha'(0)) = Df_c(v).$$

This proves $T_c M \subset \ker(Df_c)$.

**Theorem 3.7.5.** Let $U \subset \mathbb{R}^n$ be an open set and let $F : U \to \mathbb{R}^{n-k}$ be a $C^1$ mapping such that $DF_x$ is onto for all $x \in M$, so that $M = F^{-1}(0)$ is a $k$-dimensional manifold. Let $f : U \to \mathbb{R}$ be a $C^1$ function. Then $a \in M$ is a critical point of $f$ restricted to $M$ if and only if

$$\ker [DF_a] \subset T_a M \subset \ker [Df_a].$$

In turn, this is true if and only if there exist $\lambda_1, \ldots, \lambda_{n-k} \in \mathbb{R}$ such that

$$Df_a = \lambda_1 DF_1_a + \cdots + \lambda_{n-k} DF_{n-k}_a.$$
where \( F = (F_1, \ldots, F_{n-k}) \). That is, \( Df_a \in \text{span}\{D(F_1)_a, \ldots, D(F_{n-k})_a\} \), which is the subspace of \( \mathbb{R}^n \) normal (perpendicular) to \( T_aM \).

**Exercise 3.7.7.** Let \( \varphi \left( \begin{array}{c} x \\ y \\ z \\ 1 \end{array} \right) = ax + by + cz - 1 \), where \( a, b, c > 0 \).

Let \( M = \varphi^{-1}(0) \), which is a 2-dimensional manifold, a plane in fact. Let
\[
\left( \begin{array}{c} x \\ y \\ z \end{array} \right) = \left( \begin{array}{c} x \\ y \\ z \end{array} \right) _w. \]
We show that \( f \) restricted to \( M \) has four critical points.

Let \( w = \left( \begin{array}{c} x \\ y \\ z \end{array} \right) \). We compute:

\[
Df_w = \left( \begin{array}{c} yz \\ xz \\ xy \end{array} \right) \quad \text{and} \quad D\varphi_w = \left( \begin{array}{c} a \\ b \\ c \end{array} \right).
\]

By Lagrange multipliers, we have that \( w \in M \) is a critical point of \( f \) restricted to \( M \) if and only if
\[
ax + by + cz = 1
\]
and there exists \( \lambda \in \mathbb{R} \) such that
\[
Df_w = \lambda D\varphi_w.
\]
that is,

\[
\begin{align*}
yz &= \lambda a, \\
xz &= \lambda b, \\
xy &= \lambda c.
\end{align*}
\]

We have 4 equations in the 4 unknowns \(x, y, z, \lambda\). However, the equations are nonlinear and in general there is no set way to solve nonlinear equations. Fortunately, here it not hard to see how to proceed. We observe that

\[
\begin{align*}
xyz &= \lambda ax, \\
xyz &= \lambda by, \\
xyz &= \lambda cz.
\end{align*}
\]

Then

\[
\lambda = \lambda (ax + by + cz) = 3xyz.
\]

So

\[
\begin{align*}
\frac{\lambda}{3} &= \lambda ax, \\
\frac{\lambda}{3} &= \lambda by, \\
\frac{\lambda}{3} &= \lambda cz.
\end{align*}
\]

If \(\lambda = 0\), then we obtain a solution to

\[
\begin{align*}
yz &= \lambda a, \\
xz &= \lambda b, \\
xy &= \lambda c
\end{align*}
\]

if and only if two of \(x, y, z\) are zero. So the solutions are

\[
\begin{pmatrix}
\frac{1}{a} \\
0 \\
0
\end{pmatrix}, \quad \begin{pmatrix}
0 \\
\frac{1}{b} \\
0
\end{pmatrix}, \quad \begin{pmatrix}
0 \\
0 \\
\frac{1}{c}
\end{pmatrix}.
\]

On the other hand, if \(\lambda \neq 0\), then \(x = \frac{1}{3a}, y = \frac{1}{3b}, z = \frac{1}{3c}\).
To classify the critical points, we parametrize $M$ by the map $\gamma : \mathbb{R}^2 \to M$ defined by

$$\gamma(x, y) = \left(x, y, \frac{1 - ax - by}{c}\right).$$

Then

$$(f \circ \gamma)(x, y) = f\left(x, y, \frac{1 - ax - by}{c}\right) = \frac{xy - ax^2y - bxy^2}{c}$$

and it is equivalent to classify the critical points of $f \circ \gamma$. The 4 critical points of $f \circ \gamma$ are:

$$\gamma^{-1}\left(\begin{array}{c}
\frac{1}{a} \\
0 \\
0 \\
\frac{1}{c}
\end{array}\right) = \left(\begin{array}{c}
\frac{1}{a} \\
0 \\
0 \\
\frac{1}{c}
\end{array}\right), \quad \gamma^{-1}\left(\begin{array}{c}
0 \\
\frac{1}{b} \\
0 \\
\frac{1}{c}
\end{array}\right) = \left(\begin{array}{c}
0 \\
\frac{1}{b} \\
0 \\
\frac{1}{c}
\end{array}\right),$$

$$\gamma^{-1}\left(\begin{array}{c}
0 \\
0 \\
\frac{1}{a} \\
\frac{1}{c}
\end{array}\right) = \left(\begin{array}{c}
0 \\
0 \\
\frac{1}{a} \\
\frac{1}{c}
\end{array}\right), \quad \gamma^{-1}\left(\begin{array}{c}
\frac{1}{3a} \\
\frac{1}{3b} \\
\frac{1}{3c}
\end{array}\right) = \left(\begin{array}{c}
\frac{1}{3a} \\
\frac{1}{3b} \\
\frac{1}{3c}
\end{array}\right).$$

The first and second partial derivatives of $f \circ \gamma$ are:

$$D_1 (f \circ \gamma) = \frac{y - 2axy - by^2}{c},$$

$$D_2 (f \circ \gamma) = \frac{x - ax^2 - 2bxy}{c},$$

$$D_1D_1 (f \circ \gamma) = \frac{-2ay}{c},$$

$$D_2D_2 (f \circ \gamma) = \frac{-2bx}{c},$$

$$D_1D_2 (f \circ \gamma) = \frac{1 - 2ax - 2by}{c}.$$

The Hessian of $f \circ \gamma$ at $\left(\begin{array}{c}x \\
y\end{array}\right)$ is

$$[H_{ij}] = \left(\begin{array}{cc}
D_1D_1 (f \circ \gamma) & D_1D_2 (f \circ \gamma) \\
D_2D_1 (f \circ \gamma) & D_2D_2 (f \circ \gamma)
\end{array}\right) = \left(\begin{array}{cc}
\frac{-2ay}{c} & \frac{1 - 2ax - 2by}{c} \\
\frac{1 - 2ax - 2by}{c} & \frac{-2bx}{c}
\end{array}\right).$$
At the critical points we have

\[
[H_{i,j}] \begin{pmatrix} \frac{1}{a} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & -\frac{1}{c} \\ -\frac{1}{c} & -2\frac{b}{ac} \end{pmatrix}, \quad [H_{i,j}] \begin{pmatrix} 0 \\ \frac{1}{b} \end{pmatrix} = \begin{pmatrix} -2\frac{a}{bc} & -\frac{1}{c} \\ -\frac{1}{c} & 0 \end{pmatrix},
\]

\[
[H_{i,j}] \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{c} \\ \frac{1}{c} & 0 \end{pmatrix}, \quad [H_{i,j}] \begin{pmatrix} \frac{1}{3a} \\ \frac{1}{3b} \end{pmatrix} = \begin{pmatrix} -\frac{2a}{3bc} & -\frac{1}{3c} \\ -\frac{1}{3c} & -\frac{2b}{3ac} \end{pmatrix}.
\]

So

1. At \( \begin{pmatrix} \frac{1}{a} \\ 0 \end{pmatrix} \),

\[
Q_{f\circ\gamma}(h) = \frac{1}{2} \left( \begin{array}{cc} h_1 \\ h_2 \end{array} \right) \begin{pmatrix} 0 & -\frac{1}{c} \\ -\frac{1}{c} & -2\frac{b}{ac} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = -\frac{b}{ac} h_2^2 - \frac{1}{c} h_1 h_2 = \frac{1}{c} \left( -\frac{b}{a} \left( h_2 + \frac{a}{2b} h_1 \right)^2 + \frac{a}{4b} h_1^2 \right),
\]

which has signature \((1, 1)\), so we have a saddle.

2. Similarly, at \( \begin{pmatrix} 0 \\ \frac{1}{b} \end{pmatrix} \) we have a saddle. (Exercise.)

3. At \( \begin{pmatrix} 0 \\ 0 \end{pmatrix} \),

\[
Q_{f\circ\gamma}(h) = \frac{1}{2} \left( \begin{array}{cc} h_1 \\ h_2 \end{array} \right) \begin{pmatrix} 0 & \frac{1}{c} \\ \frac{1}{c} & 0 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \frac{1}{c} h_1 h_2 = \frac{1}{4c} \left( (h_1 + h_2)^2 - (h_1 - h_2)^2 \right),
\]

which has signature \((1, 1)\), so we also have a saddle.
4. At \( \left( \frac{1}{3r} \right) \),

\[
Q_{f_{\theta \gamma}}(\textbf{h}) = \frac{1}{2} \left( \begin{array}{cc}
h_1 & h_2 \\
\end{array} \right) \left( \begin{array}{cc}
\frac{-2a}{3c} & \frac{-1}{3c} \\
\frac{2b}{3ac} & \frac{1}{3c} \\
\end{array} \right) \left( \begin{array}{c}
h_1 \\
h_2 \\
\end{array} \right)
\]

\[
= -\frac{1}{3c} h_1 h_2 - \frac{1}{3} \frac{a}{b} h_1^2 - \frac{b}{3ac} h_2^2
\]

\[
= -\frac{a}{3bc} \left( h_1 + \frac{b}{2a} h_2 \right)^2 - \frac{b}{4ac} h_2^2,
\]

which is negative definite, so we have a local maximum.
Lemma 3.7.11 (Linear algebra). Let

\[ A = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix} : \mathbb{R}^n \to \mathbb{R}^m \]  
and  
\[ \beta : \mathbb{R}^n \to \mathbb{R} \]

be linear transformations. Then

\[ \ker A \subset \ker \beta \]

if and only if there exist \( \lambda_1, \ldots, \lambda_m \in \mathbb{R} \) such that

\[ \beta = \lambda_1 \alpha_1 + \cdots + \lambda_m \alpha_m. \]

Proof. (\( \Rightarrow \)) Suppose \( \beta = \lambda_1 \alpha_1 + \cdots + \lambda_m \alpha_m \) for some \( \lambda_1, \ldots, \lambda_m \in \mathbb{R} \). If \( \mathbf{v} \in \ker A \), then \( \alpha_i(\mathbf{v}) = 0 \) for all \( 1 \leq i \leq m \), so that

\[ \beta(\mathbf{v}) = \lambda_1 \alpha_1(\mathbf{v}) + \cdots + \lambda_m \alpha_m(\mathbf{v}) = 0. \]

That is, \( \mathbf{v} \in \ker \beta \). Thus \( \ker A \subset \ker \beta \).

(\( \implies \)) Suppose \( \ker A \subset \ker \beta \). Let \( \alpha_1, \ldots, \alpha_k \), where \( 1 \leq k \leq m \), be a maximal linear independent subset of the \( \alpha_i \)'s. Then

\[ \ker \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix} = \ker \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix} = \ker A. \]

Suppose \( \beta \) is not a linear combination of \( \alpha_1, \ldots, \alpha_m \), then \( \beta \) is not a linear combination of \( \alpha_1, \ldots, \alpha_k \). Hence

\[ \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \\ \beta \end{bmatrix} : \mathbb{R}^n \to \mathbb{R}^{k+1} \]

is onto. So the equation

\[ \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \\ \beta \end{bmatrix} \mathbf{v} = \mathbf{e}_{k+1} \]

has a solution. Hence \( \alpha_i(\mathbf{v}) = 0 \) for \( 1 \leq i \leq k \) and \( \beta(\mathbf{v}) = 1 \), so that \( \mathbf{v} \in \ker A \) and \( \mathbf{v} \notin \ker \beta \). This contradicts \( \ker A \subset \ker \beta \). \( \square \)