Theorem 3  

Consider the functional \( g: X \to \mathbb{R} \) on a normed linear space \( X \). Let
\[
\Omega = \{ x \in X : g_i(x) = 0, \ i = 1, 2, \ldots, n \}
\]
be a constraint set defined by the functionals \( g_i: X \to \mathbb{R}, \ i = 1, 2, \ldots, n \).

Assume \( g \) and \( g_i, \ i = 1, 2, \ldots, n \) are Fréchet differentiable.

Suppose \( x_0 \) is an extremum of \( g \) subject to the constraints \( g_i(x) = 0, \ i = 1, 2, \ldots, n \), and \( x_0 \) is a regular point of \( \Omega \). Then,
\[
\bigcap_{i=1}^{n} \ker(Dg_i(x_0)) = \ker(Dg(x_0)).
\]

Proof: Let \( h \in \bigcap_{i=1}^{n} \ker(Dg_i(x_0)) \).

By Theorem 1 and associated remark (Lect. 5A) there exist \( n \) linearly independent vectors \( y_1, y_2, \ldots, y_n \in X \) such that
\[
M = [Dg_i(x_0) y_j] = I_n \text{ the n x n}
\]

Refer to the figure on the next page depicting the constraint set as a manifold or hypersurface \( \Omega \). The plane containing \( x_0 \) is also the tangent plane to \( \Omega \) at \( x_0 \). By hypothesis, \( h \) and hence \( e h \) for any scalar \( e \) belongs to this tangent plane. (< Think about proving this>
Tangent plane to $\Omega$ at $x_0$ denoted $T_{x_0} \Omega$.

By hypothesis $h \in$ tangent plane.

But $x_0 + eh \notin \Omega$.

On the other hand the main goal of this proof is to show that there exists a suitable $y(x)$ such that

$$x_0 + eh + y(x) \in \Omega$$

provided $e$ is small enough. For this, consider the system of equations

$$g_i(x_0 + eh + \sum_{i=1}^{n} y_i x_i) = 0$$

$i = 1, 2, \ldots, n$.

where $x_0, h, y_i, i = 1, 2, \ldots, n$ are fixed as above, but $e$ and $g_i, i = 1, 2, \ldots, n$ are $(n+1)$ real valued unknowns.

Define $q = \left(\begin{array}{c} q_1 \\ \vdots \\ q_n \end{array}\right)$ and $\tilde{q} = \left(\begin{array}{c} q_1 \\ \vdots \\ q_n \end{array}\right)$. 
The above equations take the form

\[ \tilde{g} : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n \]

\[ (\varepsilon, \varphi) \mapsto \tilde{g}(\varepsilon, \varphi) = \begin{pmatrix} 
q_1(x_0 + \varepsilon h + \varphi(q_3)) \\
\vdots \\
q_n(x_0 + \varepsilon h + \varphi(q_n))
\end{pmatrix} \]

Clearly, \( \tilde{g}(0, 0) = 0 \) (since \( x_0 \in \Omega \)).

Observe that the partial Fréchet derivative

\[ D_x \tilde{g}(0, 0) = \begin{bmatrix} Dg_1(x_0) & \cdots & Dg_n(x_0) \end{bmatrix} = I_n \]

is invertible. Hence, the implicit function theorem applies at \( \varepsilon = 0, \varphi = 0 \), and there is an open interval

\[ U = (-\varepsilon_0, \varepsilon_0) \]

containing 0, such that for \( \varepsilon \in U \), there is a unique vector-valued function \( \varepsilon \mapsto \varphi(\varepsilon) \) such that

\[ \tilde{g}(\varepsilon, \varphi(\varepsilon)) = 0 \]

Then, for \( y(\varepsilon) = \sum_{i=1}^{n} \varphi_i(\varepsilon) y_i \),

\[ q_i(x_0 + \varepsilon h + y(\varepsilon)) = 0 \quad i = 1, 2, \ldots, n. \]
We let any expression \( O(\varepsilon) \) to mean that
\[
\lim_{\varepsilon \to 0} \frac{O(\varepsilon)}{\varepsilon} = 0.
\]

Also, clearly
\[
\lim_{\varepsilon \to 0} \| \varepsilon \mathbf{h} + y(\varepsilon) \| = 0 \quad \text{(recall \( y(0) = 0 \))}
\]

By definition of Fréchet derivative
\[
\lim_{\varepsilon \to 0} \frac{\| f_i(\mathbf{x}_0 + \varepsilon \mathbf{h} + y(\varepsilon)) - f_i(\mathbf{x}_0) - \nabla f_i(\mathbf{x}_0) (\varepsilon \mathbf{h} + y(\varepsilon)) \|}{\| \varepsilon \mathbf{h} + y(\varepsilon) \|} = 0
\]

equivalently, for \( i = 1, 2, \ldots, n \),
\[
f_i(\mathbf{x}_0 + \varepsilon \mathbf{h} + y(\varepsilon)) - f_i(\mathbf{x}_0) = \nabla f_i(\mathbf{x}_0) (\varepsilon \mathbf{h} + y(\varepsilon)) + o(\| \varepsilon \mathbf{h} + y(\varepsilon) \|)
\]

But the left hand side = 0 \( \text{by construction} \).

Also \( \nabla f_i(\mathbf{x}_0) \mathbf{h} = 0 \). Hence, \( y(\varepsilon) = \sum_{i=1}^{n} a_i \varepsilon^i \frac{\partial y_i}{\partial \varepsilon} \) \( \forall \varepsilon \), we note,
\[ 0 = \left[ X g_i(x, y_i) \right] \Phi(e) + o(e) + o(\|y(e)\|) \]
\[ = \Phi(e) + o(e) + o(\|y(e)\|) \]

We need the following calculations:

**dot product**

\[ y(e) = \sum_{i=1}^{n} \Phi_i(e) y_i = \sum_{i=1}^{n} (e_i \cdot \Phi(e)) y_i \]

where \( e_i = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ \vdots \end{pmatrix} \) \( i \text{th} \) place

\[ \Rightarrow \| y(e) \| \leq \sum_{i=1}^{n} |e_i \cdot \Phi(e)| \| y_i \| \]
\[ \leq \sum_{i=1}^{n} \| e_i \|_2 \| \Phi(e) \|_2 \| y_i \| \]

(Cauchy-Schwarz)

\[ = \left( \sum_{i=1}^{n} \| y_i \|^2 \right)^{\frac{1}{2}} \| \Phi(e) \|_2 \]
\[ \leq \frac{1}{d_1} \| \Phi(e) \| \leq \text{any norm, } \]

and suitable \( d_1 \)

(by equivalence of all norms in \( \mathbb{R}^n \))

\[ \Rightarrow d_1 \| y(e) \| \leq \| \Phi(e) \| \]

**ii)**

\( L : \mathbb{R}^n \rightarrow [y_1, y_2, \ldots, y_n] = \text{closed linear span of the linearly independent vectors inside square brackets (of dimension } n) \)

\( \Phi \rightarrow \sum_{i=1}^{n} \Phi_i y_i \)

is invertible (one-to-one otherwise \( y_i \), \( i = 1, 2, \ldots, n \))

(it is not a linearly independent set)
\( \varphi(x) = L^{-1} L \varphi(x) = L^{-1} y(x) \)
\[ \Rightarrow \| \varphi(x) \| = \| L^{-1} y(x) \| \]
\[ \leq \| L^{-1} \| \cdot \| y(x) \| \]
\[ = d_a \| y(x) \| \]

Combining (i) and (ii)

(iii) \( d_1 \| y(x) \| \leq \| \varphi(x) \| \leq d_2 \| y(x) \| \)

But we have shown that,

\[ 0 = \varphi(x) + 0(x) + o(\| y(x) \|) \]

Thus \( \| y(x) \| = o(x) \), using (i) and (ii).

(iii)

for \( -\varepsilon_0 < \varepsilon < \varepsilon_0 \).

Since \( x_o + \varepsilon h + y(x) \in \Omega \) for \( -\varepsilon_0 < \varepsilon < \varepsilon_0 \) it follows that

\( \Phi(x) = g(x_o + \varepsilon h + y(x)) \)

has an unconstrained extremum at 0 over the interval \( (-\varepsilon_0, \varepsilon_0) \). By Theorem 1 (Lecture 4 cont'd), it follows that

\[ 0 = \frac{d}{d\varepsilon} \Phi(x) \bigg|_{\varepsilon=0} = Dg(x_o + \varepsilon h + y(x)) \lim_{\varepsilon \to 0} \frac{\varepsilon h + y(x)}{\varepsilon} \]

\[ = Dg(x_o) h \]

\[ \Rightarrow h \in \ker(Dg(x_o)) \]