Differentiation in Linear Spaces - III

Prof. Dan A. Simovici

UMB
1. Manifolds in $\mathbb{R}^n$

2. Unconstrained Optimization

3. Optimization with Equality Constraints
Manifolds in $\mathbb{R}^n$

Definition

Let $h : X \rightarrow \mathbb{R}^m$ be a function, where $X \subseteq \mathbb{R}^n$ such that $h \in C^1(X)$. The components of $h$ are denoted by $h_1, \ldots, h_m$. A point $x_0$ that satisfies the equality $h(x_0) = 0_m$ is a regular point of $h$ if the vectors $(\nabla h_1)(x_0), \ldots, (\nabla h_m)(x_0)$ are linearly independent.
If $h : X \longrightarrow \mathbb{R}^m$, $h \in C^1(X)$ and $m \leq n$, then $x_0$ is a regular point of $h$ if the Jacobian matrix $(Dh)(x_0) \in \mathbb{R}^{m \times n}$ has rank $m$. In other words, $x_0$ is a regular point of $h$ if the vectors $(\nabla h_1)(x_0), \ldots, (\nabla h_m)(x_0)$ are linear independent.

Thus, if $h : X \longrightarrow \mathbb{R}^m$, $h \in C^1(X)$ and $m \leq n$, then $x_0$ is a regular point of $h$ if the Jacobian matrix $(Dh)(x_0) \in \mathbb{R}^{m \times n}$ has rank $m$. 
Definition

A critical point of $h$ is a point $x_0 \in X$ such that $(Dh)(x_0)$ is the zero operator, that is, $\delta f(x_0; h) = 0$ for every $h \in X$.

A critical point $x_0$ is degenerate if $H_f(x_0) = O_{n,n}$. 
Definition

A non-empty subset $M$ of $\mathbb{R}^n$ is an $r$-manifold (where $r < n$) of class $C^q(U)$ if for every $x_0 \in M$ there exists a neighborhood $U$ of $x_0$ and a function $h : U \rightarrow \mathbb{R}^{n-r}$ of class $C^q(U)$ such that the linear transformation $(Dh)(x_0)$ has rank $n - r$ and $M \cap U = \{x \in U \mid h(x) = 0_{n-r}\}$. We refer to $h$ as the locally defining function of $M$ in $x_0$.

Note that for every point $x_0$ of an $r$-manifold, the neighborhood $U$ and the locally defining function depend on $x_0$. 
$r$-manifolds determined by a function

An $r$-manifold may be defined by a single function $h$. For example, suppose that $h : \mathbb{R}^n \rightarrow \mathbb{R}^{n-r}$ is a class $C^q(X)$ function, where $X$ is an open subset of $\mathbb{R}^n$, such that $\text{rank}((Dh)(x)) = n - r$. If the set

$$M_h = \{ x \in X \mid h(x) = 0_{n-r} \}$$

is non-empty, then $M_h$ is an $r$-manifold. Indeed, it is possible to show that $\{ x \in \mathbb{R}^n \mid \text{rank}((Dh)(x)) = n - r \}$ is open. Therefore, every $x_0 \in M$ has a neighborhood $U$ such that $\text{rank}((Dh)(x)) = n - r$ for every $x \in U$ and $M_h \cap U = \{ x \in U \mid h(x) = 0_{n-r} \}$. We refer to $M_h$ as the $r$-manifold determined by $h$. 
Example

The spherical surface $S(0_n, 1)$ in $\mathbb{R}^n$ is a $(n - 1)$-manifold of class $C^2(\mathbb{R}^n)$. Indeed, let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be the function defined by $f(x) = \sum_{i=1}^n x_i^2 - 1$. We have

$$(Df)(x_0)(h) = 2x_0' h$$

so $\text{rank}((Df)(x_0)) = 1$ for every $x_0 \in U$. Also, taking $U = \mathbb{R}^n$ we have $M = \{x \in \mathbb{R}^n \mid f(x) = 0\}$. 
Example

Let $M$ be the subset of $\mathbb{R}^3$:

$$M = S(0_3, 1) \cap \{x \in \mathbb{R}^3 \mid x_1 + x_2 + x_3 = 0\}.$$

Define the function $h : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ as

$$h_1(x) = x_1^2 + x_2^2 + x_3^2 - 1,$$

$$h_2(x) = x_1 + x_2 + x_3$$

for $x \in \mathbb{R}^3$. Its Jacobian matrix is

$$(Dh)(x) = \begin{pmatrix} 2x_1 & 2x_2 & 2x_3 \\ 1 & 1 & 1 \end{pmatrix}.$$  

For any $x \in M$ we have $\text{rank}((Dh)(x)) = 2$ and

$$M = \{x \in \mathbb{R}^3 \mid h(x) = 0_2\}.\text{ Thus, } M \text{ is an 1-dimensional manifold.}$$
Example

Let 
\[ M = \{ \mathbf{x} \in \mathbb{R}^2 \mid x_1 x_2 = 1, \text{ or } x_1 = 0 \text{ and } x_2 \neq 0, \text{ or } x_2 = 0 \text{ and } x_1 \neq 0 \} \].
There are three cases to consider.
If $x_1 x_2 = 1$, define $h(x) = x_1 x_2 - 1$; we have $(Dh)(x) = (x_2, x_1)$ so in this case, $\text{rank}((Dh)(x)) = 1$. The neighborhood of $x$ is $B(x, r)$, where $0 < r < \min\{|x_1|, |x_2|\}$.

If $x_1 = 0$ and $x_2 \neq 0$, let $h_1(x) = x_1$, which implies $(Dh_1)(x) = (1, 0)$, hence $\text{rank}((Dh_1)(x)) = 1$. In this case, we can take $U = B(x, r)$, where $r < \frac{1}{|x_2|}$.

Finally, if $x_2 = 0$ and $x_1 \neq 0$, let $h_2(x) = x_2$, so $(Dh_2)(x) = (1, 0)$, hence $\text{rank}((Dh)(x)) = 1$. In this case, the neighborhood $U$ is $B(x, r)$, where $r < \frac{1}{|x_1|}$.

Thus, $M$ is an unidimensional manifold.
**Definition**

Let \( M \subseteq \mathbb{R}^n \) be a manifold of dimension \( r \), \( U \) be a neighborhood of \( x_0 \in M \), and let \( h : U \rightarrow \mathbb{R}^{n-r} \) be the locally defining function in \( x_0 \). The tangent subspace of \( M \) at \( x_0 \) is the null space \( T(x_0) \) of the linear transformation \((Dh)(x_0)\),

\[
T(x_0) = \{ t \in \mathbb{R}^n \mid (Dh)(x_0)t = 0_{n-r} \}.
\]

The subspace \((T(x_0))\perp\) is the subspace of normal vectors to \( M \) at \( x_0 \). The tangent \( r \)-plane of \( M \) at \( x_0 \) is \( \{ x_0 + t \mid t \in T(x_0) \} = t_{x_0}(T(x_0)) \).
If $M$ is an $r$-manifold such that $(Dh)(x_0)$ has rank $m = n - r$, the tangent subspace of $M$ at $x_0$ (that is, the null space of $(Dh)(x_0)$) has dimension $r$ because

$$\dim(\text{Null}((Dh)(x_0))) + \dim(\text{Img}((Dh)(x_0))) = n.$$
Example

Let

$$M = \{ x \in \mathbb{R}^3 \mid x_1^2 + x_2^2 + x_3^2 = 1 \}$$

be a 2-manifold in $\mathbb{R}^3$. For $h(x) = x_1^2 + x_2^2 + x_3^2 - 1$ we have

$$(Dh)(x) = (2x_1, 2x_2, 2x_3),$$

so $\text{rank}((Dh)(x)) = 1$. Let $a \in M$; the vector

$$(\nabla f)(a) = \begin{pmatrix} 2a_1 \\ 2a_2 \\ 2a_3 \end{pmatrix}$$

is a normal vector to $M$ in $a$. The tangent hyperplane in $a$ is

$$a_1x_1 + a_2x_2 + a_3x_3 = 0.$$
Definition

Let \( h : \mathbb{R}^n \rightarrow \mathbb{R}^{n-r} \) be a function in \( C^q(\mathbb{R}^n) \) (where \( q \geq 1 \)) and let \( g : [a, b] \rightarrow \mathbb{R}^n \) be a function such that \( h(g(t)) = 0_{n-r} \) for \( t \in [a, b] \).

The set

\[
C_g = \{ x \in \mathbb{R}^n \mid x_i = g_i(t), t \in [a, b] \}
\]

is referred to as a curve on the manifold \( M_h \).

If \( g \in C^1([a, b]) \) the curve \( C_g \) is said to be differentiable; if \( g \in C^2([a, b]) \) then \( C_g \) is twice differentiable.
Definition

Let $X$ be a open subset in $\mathbb{R}^n$ and let $f : X \rightarrow \mathbb{R}$ be a functional. The point $x_0 \in X$ is a local minimum for $f$ if there exists $\delta > 0$ such that $B(x_0, \delta) \subseteq X$ and $f(x_0) \leq f(x)$ for every $x \in B(x_0, \delta)$.

The point $x_0$ is a strict local minimum if $f(x_0) < f(x)$ for every $x \in B(x_0, \delta) - \{x_0\}$. 
The notions of local maximum and strict local maximum are defined similarly. Namely, \( x_0 \in X \) is a **local maximum** for \( f \) if there exists \( \delta > 0 \) such that \( B(x_0, \delta) \subseteq X \) and \( f(x_0) \geq f(x) \) for every \( x \in B(x_0, \delta) \). The point \( x_0 \) is a **strict local maximum** if \( f(x_0) > f(x) \) for every \( x \in B(x_0, \delta) - \{x_0\} \). A **local extremum** of a functional \( f : X \rightarrow \mathbb{R} \) is a local maximum or a local minimum of \( f \).
Theorem

Let $X$ be a open subset in $\mathbb{R}^n$ and let $f : X \rightarrow \mathbb{R}$ be a functional that has a Gâteaux derivative on $X$. Every local extremum point of $f$ is a critical point of $f$. 
Define the function $g : \mathbb{R} \rightarrow \mathbb{R}$ as $g(a) = f(x_0 + ah)$. Since $x_0$ is a local extremum the function $g$ is differentiable and has a minimum in $a = 0$. Therefore, $g'(0) = 0$ and

$$
g'(0) = \lim_{r \to 0} \frac{g(r) - g(0)}{r} = \lim_{r \to 0} \frac{f(x_0 + rh) - f(x_0)}{r} = \left( (Df)(x_0) \right)' h = \delta f(x_0; h) = 0,
$$

for every $h \in X$. 
Definition

Let $U$ be a open subset in $\mathbb{R}^n$. If $x_0 \in U$, then $d$ is a feasible direction at $x_0$ for $U$ if $d \neq 0_n$ and there exists $a > 0$ such that $x_0 + td \in U$ for every $t \in [0, a]$.

If $x_0 \in \text{I}(U)$, then there exists an open sphere $B(x_0, r)$ included in $U$, so every direction $d$ is feasible. The set of feasible directions at $x_0$ is denoted by $\text{FD}(U, x_0)$.
Definition

Let $U$ be a open subset in $\mathbb{R}^n$ and let $f : U \rightarrow \mathbb{R}$ be a functional. If $x_0 \in K(U)$ then $d \in S$ is a descent direction for $f$ at $x_0$ if there exists $\delta > 0$ such that $f(x_0 + ad) < f(x_0)$ for $0 < a < \delta$.

d is an ascent direction of $f$ at $x_0$ if if there exists $\delta > 0$ such that $f(x_0 + ad) > f(x_0)$ for $0 < a < \delta$. 
Proof

We give the argument for the case when \(((\nabla f)(x_0))'d < 0\). Since \(f\) is Gâteaux differentiable at \(x_0\) we have

\[
\begin{align*}
    f(x_0 + ad) - f(x_0) &= a((\nabla f)(x_0))'(d) = o(d),
\end{align*}
\]

Therefore,

\[
\begin{align*}
    f(x_0 + ad) - f(x_0) &= a((\nabla f)(x_0))'d + o(d).
\end{align*}
\]

Since \(\lim_{d \to 0} o(d) = 0\), for a sufficiently small \(a\) we have

\[
\begin{align*}
    f(x_0 + ad) - f(x_0) &< 0,
\end{align*}
\]

so \(d\) is indeed a descent direction at \(x_0\).
Theorem

Let $U$ be an open subset of $\mathbb{R}^n$ and let $f : U \rightarrow \mathbb{R}$ be a functional that has a Gâteaux derivative on $U$. If $x_0$ is a local minimum (maximum) of $f$, then every feasible direction at $x_0$ is an ascent (descent) direction.
Suppose that $x_0$ is a local minimum of $f$ and that $d$ is a feasible direction at $x_0$, that is, $x_0 + td \in U$ for $t \in [0, a]$ (for some $a > 0$). Let $g : [0, a] \rightarrow \mathbb{R}$ be the function defined by $g(t) = f(x_0 + td)$. Clearly, $g$ has a local minimum at $t = 0$, so $g'(0) \geq 0$. Since $\frac{d}{dt} g(0) = (\nabla f)(x_0)'d \geq 0$, it follows that $d$ is an ascent direction.
The set of descent directions of a functional $f : U \rightarrow \mathbb{R}$ at $x_0$ that is Gâteaux differentiable is denoted by:

$$F_0(f, x_0) = \{ r \in \mathbb{R}^n \mid (\nabla f)(x_0)'r < 0 \}$$

If $x_0$ is a local minimum for $f$, then

$$FD(f, x_0) \cap F_0(f, x_0) = \emptyset.$$
Theorem

Let \( f : B(x_0, r) \rightarrow \mathbb{R} \) be a function that belongs to the class \( C^2(B(x_0, r)) \), where \( B(x_0, r) \subseteq \mathbb{R}^k \) and \( x_0 \) is a critical point for \( f \). If the Hessian matrix \( H_f(x_0) \) is positive semidefinite, then \( x_0 \) is a local minimum for \( f \); if \( H_f(x_0) \) is negative semidefinite, then \( x_0 \) is a local maximum for \( f \).
Proof

The Taylor formula implies

\[ f(x) = f(x_0) + \left( (h' \nabla f)(x_0) + h' H_f(x_0 + \theta h)h \right), \]

where \( h = x - x_0 \) is such that \( \| h \| \leq r \). Since \( x_0 \) is a critical point, it follows that

\[ f(x) = f(x_0) + h' H_f(x_0 + \theta h)h. \]

Therefore, if \( H_f(x_0) \) is positive semidefinite we have \( h' H_f(x_0)h \geq 0 \) for \( h \in \mathbb{R}^k \). Since the second derivatives of \( f \) are continuous, if \( \theta \) is sufficiently small, \( H_f(x_0 + \theta h) \) is also positive semidefinite, hence \( f(x) \geq f(x_0) \), which means that \( x_0 \) is a local minimum; if \( H_f(x_0) \) is negative semidefinite, it follows that \( f(x) \leq f(x_0) \) so \( x_0 \) is a local maximum for \( f \).
If $x_0$ is a non-degenerate critical point for $f$ and $H_f(x_0)$ is neither positive semidefinite nor negative semidefinite, then $x_0$ is a saddle point for $f$. 
Example

Let \( f : \mathbb{R}^2 \rightarrow \mathbb{R} \) be a function in \( C^2(B(x_0, r)) \). The Hessian matrix in \( x_0 \) is

\[
H_f(x_0) = \begin{pmatrix}
\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\
\frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2}
\end{pmatrix}(x_0).
\]

Let \( a_{11} = \frac{\partial^2 f}{\partial x_1^2}(x_0) \), \( a_{12} = \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_0) \), and \( a_{22} = \frac{\partial^2 f}{\partial x_2^2}(x_0) \).
Note that

\[ h' H_f(x_0)h = a_{11} h_1^2 + 2a_{12} h_1 h_2 + a_{22} h_2^2 \]

\[ = h_2^2 (a_{11} \xi^2 + 2a_{12} \xi + a_{22}) , \]

where \( \xi = \frac{h_1}{h_2} \). For a critical point \( x_0 \) we have:

1. \( h' H_f(x_0)h \geq 0 \) for every \( h \) if \( a_{11} > 0 \) and \( a_{12}^2 - a_{11} a_{22} < 0 \); in this case, \( H_f(x_0) \) is positive semidefinite and \( x_0 \) is a local minimum;

2. \( h' H_f(x_0)h \leq 0 \) for every \( h \) if \( a_{11} < 0 \) and \( a_{12}^2 - a_{11} a_{22} < 0 \); in this case, \( H_f(x_0) \) is negative semidefinite and \( x_0 \) is a local maximum;

3. if \( a_{12}^2 - a_{11} a_{22} \geq 0 \); in this case, \( H_f(x_0) \) is neither positive nor negative definite, so \( x_0 \) is a saddle point.
Note that in the first two previous cases we have $a_{12}^2 < a_{11}a_{22}$, so $a_{11}$ and $a_{22}$ have the same sign.
Example

Let $a_1, \ldots, a_m$ be $m$ points in $\mathbb{R}^n$. The function $f(x) = \sum_{i=1}^{m} \| x - a_i \|^2$ gives the sum of the squares of the distances between $x$ and the points $a_1, \ldots, a_m$. We will prove that this sum has a global minimum obtained when $x$ is the barycenter of the set $\{a_1, \ldots, a_m\}$. 
Example (cont’d)

We have:

\[ f(x) = m \| x \|^2 - 2a'_i x + \| a_i \|^2 \]

\[ = m(x_1^2 + \cdots + x_n^2) - 2 \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij} x_j + \| a_i \|^2, \]

which implies

\[ \frac{\partial f}{\partial x_j} = 2mx_j - 2 \sum_{i=1}^{m} a_{ij} \]

for \( 1 \leq j \leq n \). Thus, there exists only one critical point given by

\[ x_j = \frac{1}{m} \sum_{i=1}^{m} a_{ij} \]

for \( 1 \leq j \leq n \). The Hessian matrix \( H_f = 2ml_n \) is positive definite, so the critical point is a local minimum and, in view of convexity of \( f \), the global minimum. This point is the barycenter of the set \( \{a_1, \ldots, a_m\} \).
**Example**

Let \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) be a linear functional defined as \( f(x) = a'x + b \) whose values \( y_1, \ldots, y_m \) are obtained by performing a series of experiments starting with the values \( x_1, \ldots, x_m \in \mathbb{R}^n \) of the input parameters of the experiments. The goal of the experiments is to determine \( a \in \mathbb{R}^n \) and \( b \in \mathbb{R} \). Noise and experimental errors affect the values of the results of the experiment such that the system in \( a \) and \( b \):

\[
a'x_k + b = y_k
\]

for \( 1 \leq k \leq m \) is not compatible in general. The next best thing to solving this system is to determine \( a \) and \( b \) such that the square error

\[
r(a_1, \ldots, a_n, b) = \sum_{k=1}^{m} (a'x_k + b - y_k)^2
\]

is minimal.
Unconstrained Optimization

Example (cont’d)

If \( X = (1_n \ x_1 \cdots x_m) \in \mathbb{R}^{n \times (1+m)} \) and \( y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \), then \( r \) can be written as:

\[
 r(c) = \| y - Xc \|^2,
\]

where \( c = \begin{pmatrix} b \\ a \end{pmatrix} \). Observe that

\[
 r(c) = (y - Xc)'(y - Xc) = y'y - y'Xc - c'X'y + c'X'Xc,
\]

where \( c'X'y = y'Xc \) is a scalar (and, therefore, equals its transpose). Thus,

\[
 r(c) = y'y - 2y'Xc + c'X'Xc.
\]

The necessary conditions for the minimum yield \((D_c r)(c_0) = 0\), which amount to

\[
 -X'y + (X'X)c = 0.
\]
**Example**

Let \( f : \mathbb{R}^2 \rightarrow \mathbb{R} \) be the function defined by \( f(x) = \sin(x_1 x_2) \). We have

\[
\frac{\partial f}{\partial x_1} = x_2 \cos(x_1 x_2) \quad \text{and} \quad \frac{\partial f}{\partial x_2} = x_1 \cos(x_1 x_2),
\]

which yield

\[
\frac{\partial^2 f}{\partial x_1^2} = x_2^2 \cos(x_1 x_2), \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = \cos(x_1 x_2) - x_1 x_2 \sin(x_1 x_2), \quad \frac{\partial^2 f}{\partial x_2^2} = x_1^2 \cos x_1 x_2.
\]

The set of critical points of \( f \) are given by \( x_1 x_2 = \pi(k \pm \frac{\pi}{2}) \) for \( k \in \mathbb{Z} \). On all these points we have

\[
\frac{\partial^2 f}{\partial x_1^2}(x) = \frac{\partial^2 f}{\partial x_2^2}(x) = 0,
\]

so all these critical points are saddle points.
Theorem

(Existence Theorem of Lagrange Multipliers) Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and let $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be two functions such that

1. $f \in C^1(\mathbb{R}^n)$ and $h \in C^1(\mathbb{R}^n)$, and
2. $\text{rank}((Dh)(x)) < n$.

If $x_0$ is a regular point of $h$ and a local extremum of $f$ subjected to the restriction $h(x_0) = 0_m$, then $(\nabla f)(x_0)$ is a linear combination of $(\nabla h_1)(x_0), \ldots, (\nabla h_m)(x_0)$. 
Proof

A tangent vector $y$ to the manifold $M_h$ can be obtained starting from a smooth curve $C_g$ contained in $M_h$ that passes through $x_0$. In this case, there exist $a, b$ in $\mathbb{R}$ such that $g : (a, b) \rightarrow \mathbb{R}^n$ is a function such that $h(g(t)) = 0_{n-r}$ for $t \in (a, b)$, and $x_0 = g(t_0)$ for some $t_0 \in (a, b)$.

Let $\phi : [a, b] \rightarrow \mathbb{R}^n$ be defined by $\phi(t) = f(g(t))$ for $t \in [a, b]$.

Since $x_0$ is a local maximum for $f$, $t_0$ is a local maximum for $\phi$ and $\frac{d}{dt}\phi(t_0) = 0$. The chain rule implies:

$$\frac{d\phi}{dt}(t_0) = (\nabla f)(g(t_0))(Dg)(t_0) = (\nabla f)(g(t_0))y.$$
Thus, \((\nabla f)(x_0)\) is perpendicular on every vector tangent to \(M_h\), so it is a normal vector on \(M_h\) at \(x_0\). Since \((\nabla h_1)(x_0), \ldots, (\nabla h_m)(x_0)\) is a basis of the subspace normal vectors at \(x_0\), it follows that \((\nabla f)(x_0)\) is a linear combination of \((\nabla h_1)(x_0), \ldots, (\nabla h_m)(x_0)\). We can write

\[
(\nabla f)(x_0) = \lambda_1(\nabla h_1)(x_0) + \cdots + \lambda_m(\nabla h_m)(x_0).
\]

The coefficients \(\lambda_1, \ldots, \lambda_m\) are known as Lagrange multipliers.
Let us consider a variant of the optimization problem previously discussed. As before, we start with $m$ points in $\mathbb{R}^n$, $a_1, \ldots, a_m$. This time we seek to minimize $f(x) = \sum_{i=1}^{m} \| x - a_i \|^2$ subjected to the restriction $\| x - b \| = r$, where $b \in \mathbb{R}^n$ and $r \geq 0$. Equivalently, this restriction is equivalent to $\| x - b \|^2 = r^2$.

We saw that

$$\frac{\partial f}{\partial x_j} = 2mx_j - 2 \sum_{i=1}^{m} a_{ij}$$

for $1 \leq j \leq n$, so

$$(\nabla f)(x_0) = 2mx - 2A1_m,$$

where $A = (a_1, \ldots, a_m) \in \mathbb{R}^{n \times m}$. 
Example (cont’d)

There exists $\lambda$ such that

$$2m\mathbf{x}_0 - 2A\mathbf{1}_m = 2\lambda(\mathbf{x}_0 - \mathbf{b}),$$

hence $\mathbf{x}_0$ satisfies the equality

$$\mathbf{x}_0 = \frac{1}{m - \lambda} (A\mathbf{1}_m - \lambda\mathbf{b}).$$

Since $\mathbf{x}_0$ must be located on the sphere $S(\mathbf{b}, r)$ it follows that we must have

$$\lambda = m - \frac{1}{r} \| A\mathbf{1}_m - m\mathbf{b} \|,$$

which means that the extremum is reached when

$$\mathbf{x}_0 = r \frac{A\mathbf{1}_m - m\mathbf{b}}{\| A\mathbf{1}_m - m\mathbf{b} \|}.$$
If \( \mathbf{c} \) is the barycenter of the set \( \{\mathbf{a}_1, \ldots, \mathbf{m}\} \), \( \mathbf{c} = \frac{1}{m} \, A1_m \), the extremum can be written as

\[
\mathbf{x}_0 = r \, \frac{\mathbf{c} - \mathbf{b}}{\| \mathbf{c} - \mathbf{b} \|}.
\]