10 Constrained Optimization

10.1

Lemma 10.1

Let $f \in C^1(\Omega)$ and $\Omega \subset \mathbb{R}^N$, $S \subset \Omega$
Assume $x \in S$ and $f(x) = \min_{\tilde{x} \in S} f(\tilde{x})$

Then, $\langle \nabla f(x), v \rangle \geq 0$ forall $v \in T_x S$

(Also writes $D_x f\, v \geq 0$, forall $v \in T_x S$.)

proof. Let $(x_n) \subset S$ with $x_n \to x$,
$(\lambda_n) \subset (0,1)$ with $\lambda_n \to 0$,

and $\frac{x_n - x}{\lambda_n} \to v$

We already know $\frac{f(x_n) - f(x)}{\lambda_n} \to D_x f\, v$

Assume $D_x f\, v < 0$, for a contradiction, for some $v \in T_x S$.

Let $D_x f\, v = -\varepsilon < 0$ for some $\varepsilon > 0$

Let $N \in \mathbb{N}$ be such that forall $n \geq N$,

$$\left| \frac{f(x_n) - f(x)}{\lambda_n} - D_x f\, v \right| < \frac{1}{2}\varepsilon$$

We see that $\frac{f(x_N) - f(x)}{\lambda_N} < -\frac{1}{2}\varepsilon$

but $\lambda_N > 0$, hence
$f(x_N) < -\frac{1}{2}\varepsilon \cdot \lambda_N + f(x) < f(x)$ ↯ $\square$

corollary: if S is open, then

$$\langle \nabla f(x), v \rangle = 0, \quad \text{forall } v \in T_x S = \mathbb{R}^N$$

proof: let $v \in \mathbb{R}^N = T_x S$.
then

$$\langle \nabla f(x), v \rangle \geq 0 .$$

But also $-v \in \mathbb{R}^N$, and

$$0 \leq \langle \nabla f(x), -v \rangle = -\langle \nabla f(x), v \rangle \leq -0 = 0$$

Therefore, $\langle \nabla f(x), v \rangle = 0$ forall $v \in \mathbb{R}^N$

$\square$

10.2

definition 10.4 (Manifold)

Let $S \subset \mathbb{R}^N$ and $x \in S$. We say that S is a submanifold of class $C^1(U)$ and of dimension $N-k$
if

There exists $U \subset \mathbb{R}^N$ open and $x \in U$ and $C^1(U)$ functions $\Phi_1 .. \Phi_k : U \to \mathbb{R}$ such that $\{\nabla \phi_1(\tilde{x}) .. \nabla \phi_k(\tilde{x})\}$ lin. indep, $\forall \tilde{x} \in U$

$$S \cap U = \{\tilde{x} \in U \mid \Phi_1(x) = \dots = \Phi_k(x) = 0\}$$

In words, there is a neighbourhood U, open, of x such that inside U, S is the solution set of k C^1 equations, and $\operatorname{rank}(D_{\tilde{x}} \Phi) = k$, forall $\tilde{x} \in U$

By the implicit function theorem, we have a different definition involving a local parametrization, too:

proposition 10.5

i) Let $S \subset \mathbb{R}^N$ be an $N-k$ dimensional submanifold at $x \in S$.

ii) This is equivalent to: we can find a $U \subset \mathbb{R}^N$ open with $U \ni x$ and a $C^1(U)$ diffeomorphism

$$\varphi : U \to \mathbb{R}^N \quad \text{such that} \quad \Phi(S \cap U) = A \times \{0\}$$

where $A \subset \mathbb{R}^{N-k}$ is open and $0 \in \mathbb{R}^k$

proof:

$\Rightarrow$ assume i) and let $U \subset \mathbb{R}^N$ be such that $x \in U$ and U open and there exists

$$\Phi \in C^1(U, \mathbb{R}^k) \quad \text{with} \quad S \cap U = \{\tilde{x} \in U \mid \Phi(\tilde{x}) = 0\}$$
$$\text{and} \quad \operatorname{rank}(D_{\tilde{x}} \Phi) = k \quad , \forall \tilde{x} \in U$$

If we write $\Phi : \mathbb{R}^{N-k} \times \mathbb{R}^k \to \mathbb{R}^k$, $(x,y) \mapsto \Phi(x,y)$ and by $\operatorname{rank}(D_x \Phi) = k$, assume wlog

that the last k columns of $J_x \Phi$ are linearly indep. in $\mathbb{R}^k$,

then $\det\left(\frac{\partial \Phi}{\partial y}\right)(x',y) \neq 0$, where $x = (x', y)$

therefore, we can find open $X \subset \mathbb{R}^{N-k}$, $Y \subset \mathbb{R}^k$ with $x \in X \times Y$ such that there is a $C^1(X,Y) \ni g$ with

$$\{\Phi = 0\} \cap (X \times Y) = \{(x', g(x')) \mid x' \in X\}$$

then, define a function $\varphi : X \times Y \to \mathbb{R}^N$ through

$$\varphi(x', y) = (x', \Phi(x', y))$$

Then, we obviously still have $\varphi(U \cap S) = X \times \{0\}$ where $X \subset \mathbb{R}^{N-k}$ is open.

and further,

$$\frac{\partial \varphi}{\partial x}(x) = \left(\begin{array}{c|c} Id_{N-k} & 0 \\ \hline \frac{\partial \Phi}{\partial x'}(x) & \frac{\partial \Phi}{\partial y}(x) \end{array} \right) = \frac{\partial \Phi}{\partial y}(x) \neq 0$$

So by the inverse function theorem, φ is locally on some $x \in A \subset X$ open, a diffeomorphism with for $U = A \times Y$,

$$\varphi(U \cap S) = A \times \{0\}.$$

$\Leftarrow$ Let S be locally described at x as

$$S := \{ x \in U \mid \varphi(x) \in A \times \{0\} \},$$

where $\varphi \in C^1(U, \mathbb{R}^N)$ is a <u>diffeomorphism</u> and $A \subset \mathbb{R}^{N-k}$ is an open set.

Let Φ_i be the $(N-k+i)$th component of φ, so $\varphi = (\varphi_1 \ldots \varphi_{N-k}, \Phi_1 \ldots \Phi_k)$

Then Φ_i is $C^1(U)$ because φ is, and

$$\Phi(\tilde{x}) = (\Phi_1 \ldots \Phi_k)(\tilde{x}) = 0, \quad \forall \tilde{x} \in S \cap U$$

This gives $\Phi(S \cap U) \subset \{0\}$, hence

$$x \in S \cap U \iff \Phi(x) = 0$$

to show $\nabla\Phi_1(x) \,..\, \nabla\Phi_k(x)$ are lin. indep. in a local neighbourhood $V \subset \mathbb{R}^N$ of $\tilde{x}$, it suffices to note

$$D_{\tilde{x}}\varphi = \begin{pmatrix} \nabla\varphi_1(x) \\ \vdots \\ \nabla\varphi_{N-k}(x) \\ \nabla\Phi_1(x) \\ \vdots \\ \nabla\Phi_k(x) \end{pmatrix}$$

has full rank (is invertible)

So we must have, by continuity of $x \mapsto \nabla\Phi_i(x)$, $i = 1..k$, that for $\tilde{x} \in V \subset \mathbb{R}^N$ open with $\tilde{x} \in V$, we have

$$\{\nabla\Phi_1(\tilde{x}) \;\; \nabla\Phi_k(\tilde{x})\}$$ linearly independent.

$\square$

proposition 10.7

Let $S \subset \mathbb{R}^N$ be an $N-k$-dim. submanifold at x and let $U \subset \mathbb{R}^N$ be the open nbrhood from def. and $(\Phi_1 .. \Phi_k) = \Phi \in C^1(U, \mathbb{R}^k)$ be from def. s.t

$$S \cap U = \{\Phi \equiv 0\}$$

Then,

$$T_xS = \ker(D_x\Phi)$$

This makes T_xS a linear space of dimension $N - \operatorname{rank} D_x\Phi$
$= N - k$

proof

$\subset$ Let $v \in T_x S$, meaning there is a
$(x_n) \subset S$, $x_n \to x$
$(\lambda_n) \subset (0,1)$, $\lambda_n \to 0$ such that $\frac{x_n - x}{\lambda_n} \to v$

Up to considering $n \geq N$ for $N \in \mathbb{N}$ large enough that $x_n \in S$, we assume $(x_n) \subset S$.

We already have from Ch. 6 that

$$\lim_{n\to\infty} \frac{\Phi_i(x_n) - \Phi(x)}{\lambda_n} = \langle \nabla \Phi_i(x), v \rangle$$

And by $(x_n) \subset S$, this means $\Phi_i(x_n) = \Phi_i(x) = 0$ forall $n \in \mathbb{N}$. So $\langle \nabla \Phi_i(x), v \rangle = 0 \quad \forall i = 1..k$.

$\supset$ To prove every $v \in \ker(D_x \Phi)$ is tangent to S at x, we argue that $\ker(D_x \Phi)$ is a lin. space of dim. $N-k$, so it is enough to show $T_x S$ is, too.

Use that there is a $V \subset \mathbb{R}^N$ open, $x \in V$ and a diffeomorphism $\varphi : V \to \mathbb{R}^N$ with

$$\varphi(V \cap S) = A \times \{0\}$$

where $0 \in \mathbb{R}^k$, $A \subset \mathbb{R}^{N-k}$ (Proposition 10.5)

Since φ is a diffeomorphism, forall $\tilde{x} \in V$ and all $Z \subset V$ it holds

$$T_{\varphi\tilde{x}}\, \varphi(Z) = (D_{\tilde{x}} \varphi)[T_{\tilde{x}} Z]$$

in particular, $T_{\varphi(x)}[A \times \{0\}] = (D_x\varphi)[T_x(S \cap V)]$

and since V is open, $T_x(S \cap V) = T_x S$

Clearly, since $A \subset \mathbb{R}^{N-k}$ is open, it contains

$$\varphi(x) + \delta_i \cdot e_i \quad , \; i = 1 .. N-k$$

for $\delta_i > 0$ sufficiently small.
Therefore, if we let $y_n^i = \varphi(x) + \frac{1}{n} e_i$
for $n \geq N_i$ s.t $\frac{1}{n} < \delta_i$, we see

$$\frac{1}{n} \to 0, \quad \varphi(x) + \frac{1}{n} e_i \to 0,$$

$$\text{and} \quad \frac{\varphi(x) + \frac{1}{n} e_i - \varphi(x)}{\frac{1}{n}} \to e_i$$

hence

$$T_{\varphi(x)}(A \times \{0\}) \ni e_i \, , \; i = 1, N-k$$

More generally, for any $v \in \mathbb{R}^{N-k} \times \{0\}$
we can pick $y_n^v = \varphi(x) + \frac{1}{n} v$
which is $\in A \times \{0\}$ for $n \geq N$ suff. large,

So $\mathbb{R}^{N-k} \times \{0\} \subset T_{\varphi(x)}(A \times \{0\})$

We also see that if $v \in T_{\varphi(x)}(A \times \{0\})$,
then

$$v_j = 0 \quad \text{for} \quad j > N-k,$$

since we have $[y_n]_j = 0$ for any
sequence $(y_n) \subset A \times \{0\}$ and $\varphi_j(x) = 0$ as well.

Therefore, $\mathbb{R}^{N-k} \times \{0\} \supset T_{\varphi(x)}(A \times \{0\})$, and we conclude

$$(D_x\varphi)(T_xS) = \mathbb{R}^{N-k} \times \{0\}$$

By $\det(D_x\varphi) \neq 0$, it follows that T_xS is a $N-k$ dim. lin. space and it contains $\ker(D_x\varphi)$ with dim $N-k$

conclusion: $\ker(D_x\varphi) = T_xS$ $\square$

Example: consider $\mathbb{R}^{n\times n}$ as an n^2-dimensional vector space with basis

$$\beta = \{ B^{(ij)} \mid i=1..n, j=1..n \}$$

Where $B^{(ij)}_{k\ell} = \delta_{ik}\delta_{j\ell}$ (Kronecker delta)

Let $S = SL(n,\mathbb{R})$, the special linear group of $M \in \mathbb{R}^{n\times n}$ with $\det(M) - 1 = 0$

Since $\det : \mathbb{R}^{n\times n} \to \mathbb{R}$ is polynomial in the entries M_{ij}, it is $C^1(\mathbb{R}^{n\times n})$,

$$\Phi(M) := \det(M) - 1$$

Is $C^1(\_\ \mathbb{R}^{n\times n})$

This means that if we can show $\nabla\Phi(M) \neq 0$ for any $M \in SL(\mathbb{R},n)$, then $SL(\mathbb{R},n)$ is a manifold of dim. n^2-1 at any of its elements.

to show this, note $\partial_B \Phi(M) = \partial_B \det(M)$

and we already showed in HW 5, Ex. 2 that

$$\det(I_n + tB) = 1 + t\,\mathrm{tr}(B) + o(t)$$

So
$$\det(M + tB) = \det(I_n + tM^{-1}B)\det(M)$$
$$= \det(M) + t\det(M)\mathrm{tr}(M^{-1}B) + o(t)$$

Hence $\partial_B \det(M) = \det(M)\,\mathrm{tr}(M^{-1}B) \neq 0$, for at least one $B \in \beta$ and we conclude $SL(\mathbb{R}, n)$ is a manifold at any $M \in SL(\mathbb{R}, n)$.

theorem 10.2 (Lagrange Multipliers)

Let $f : \Omega \to \mathbb{R}$ be of $C^1(\Omega)$ and $\Omega \subset \mathbb{R}^N$ open
Let $\Phi : \Omega \to \mathbb{R}^k$ be of $C^1(\Omega)$ and

$$S := \{ x \in \Omega : \Phi(x) = 0 \}$$

Assume x is a point of (local) minimum over S,

$$f(x) = \min_{y \in S} f(y)$$

(if only local, make Ω smaller!)

Assume that $\{ \nabla\Phi_1(x) \dots \nabla\Phi_k(x) \}$ is lin. indep.
I.e. S is a manifold at x

Then
$$\nabla f(x) = \sum_{i=1}^{k} \lambda_i \nabla\Phi_i(x),$$

for existent $\lambda_1 \dots \lambda_k \in \mathbb{R}$

proof: We already know that forall $v \in T_x S$,

$$\langle \nabla f(x), v \rangle \geq 0$$

since $T_x S$ is a linear subspace, also $-v \in T_x S$ forall $v \in T_x S$, so we have

$$\text{forall } v \in T_x S: \langle \nabla f(x), v \rangle = 0$$

moreover, $T_x S = \left(\operatorname{span}\{\nabla \Phi_i(x)\}_{i=1}^k\right)^\perp$

While $\langle \nabla f(x), v \rangle = 0$ means $v \in \left(\operatorname{span}\{\nabla f(x)\}\right)^\perp$

So, we have $\left(\operatorname{span}\{\nabla \Phi_i(x)\}_{i=1}^k\right)^\perp \subset \left(\operatorname{span}\{\nabla f(x)\}\right)^\perp$

for linear spaces. We already know that if $V, W \subset X$ are subspaces of a linear space X, then

$$V^{\perp\perp} = V \quad \text{and} \quad V^\perp \subset W^\perp \Leftrightarrow V \supset W$$

Therefore, $\{\nabla f(x)\} \subset \operatorname{span}\{\nabla f(x)\} \subset \operatorname{span}\{\nabla \Phi_i(x)\}_{i=1}^k$

So it follows $\nabla f(x) = \sum_{i=1}^k \lambda_i \nabla \Phi_i(x)$ for (unique) scalars $\lambda_1 \ldots \lambda_k$

$\square$

Existence of a minimum is not proved by this theorem: this should be shown separately, for example via compactness of S and continuity of f (Weierstraß)

The Lagrange Multiplier theorem gives a system of $N+k$ equations in $N+k$ unknowns, namely:

$$\begin{cases} \left(\Phi_1(x) \ \ldots \ \Phi_k(x)\right) = 0 & k \text{ eq.} \\ \nabla f(x) = \sum_{i=1}^k \lambda_i \nabla \Phi_i(x) & N \text{ eq.} \end{cases}$$

$\lambda_1 \ldots \lambda_k, x_1 \ldots x_N$ $\qquad$ $N+k$ unknowns

Note : if $\nabla f(x) = 0$, then the λ_i are all 0 by linear independence of the $\nabla \Phi_i(x)$.

Geometric interpretation : locally around $x \in S$, we have for any $v \in S^{N-1}$:

$$\langle \nabla f(x), v \rangle \leq |\nabla f(x)|$$

and this is an equality for $v = \frac{\nabla f(x)}{|\nabla f(x)|}$

In $S = \{ \Phi_i \equiv 0 \, ; i = 1..k \}$ we have

$$T_x S = \left(\text{span} \{ \nabla \Phi_i(x) \}_{i=1}^{k} \right)^{\perp}$$

So if $\nabla f(x)$ and $\{ \nabla \Phi_i(x) \}_{i=1}^{k}$ were not linearly dependent,

$$\text{span} \{ \nabla f(x), \nabla \Phi_1(x) .. \nabla \Phi_k(x) \} \supsetneq \text{span} \{ \nabla \Phi_1(x), .. , \nabla \Phi_k(x) \}$$

and in particular, we can find a $v \in S^{N-1}$ with

$$\langle \nabla \Phi_i(x), v \rangle = 0 \quad \text{forall } i = 1 .. k$$
$$\langle \nabla f(x), v \rangle \neq 0$$

meaning we can locally ~~in~~decrease along v, a contradiction.

Thus, $\nabla \Phi_i(x)$ can be seen as the "normal vectors" to S at x. If S is open, it has no normal vectors and therefore $\nabla f(x) = 0$ necessarily. Another way to understand this is that we can "walk in any direction v" and $\langle \nabla f(x), v \rangle$ has to be 0 for all such directions.

10.3 **Theorem 10.20** Mixed constraints:
Karush – Kuhn – Tucker conditions.

Let $\Omega \subset \mathbb{R}^N$ open and $f : \Omega \to \mathbb{R}$ be $C^1(\Omega)$

Let $E \subset \Omega$ be defined as

$$E = \{ x \in \Omega \mid \Phi_i(x) \overset{(1)}{\leq} 0 \ , \text{forall } i = 1..k \}$$

Where each $\Phi_i \in C^1(\Omega)$.
Let $x \in E$ be a point of local minimum of f on E

Let $r \in \{1..k\}$ be such that

(1) is strict: $\Phi_1(x) < 0 \ldots \Phi_r(x) < 0$

(1) is an equality: $\Phi_{r+1}(x) = 0 \ldots \Phi_k(x) = 0$

And $\{ \nabla\Phi_{r+1}(x), \ldots \nabla\Phi_k(x) \}$ is lin. indep.

Then

$$T_x E = \{ v \in \mathbb{R}^N \mid \langle v, \nabla\Phi_i(x) \rangle \leq 0, \text{ forall } i = r+1 \ldots k \}$$

And in particular, $\langle \nabla f(x), v \rangle \geq 0$, forall $v \in T_x E$

Note: if there is no such r, then E is clearly open, therefore $T_x E = \mathbb{R}^N$, which does not contradict the theorem

Note: if $r = 0$, this is the Lagrange multiplier theorem. and we have both Φ_i and $-\Phi_i$ as $\leq$.

proof.

We use the following lemma:

10.22. If $v_1 .. v_k \in V$, where V is an inner product space, are linearly independent, then there is a $w \in V$ with $\langle v_i, w \rangle < 0 \quad \forall i = 1..k$.

and even $w \in \operatorname{span}\{v_1 .. v_k\}$

proof: use Gram-Schmidt to find an orthogonal basis $\{u_1 .. u_k\}$ for $\operatorname{span}\{v_1 .. v_n\}$

Let $v_i = \sum_{j=1}^{k} A_{ij} u_j$

Let

$$w = \sum_{j=1}^{k} B_j u_j$$

We need to find $B_1, .., B_k$ with

$$\langle v_i, w \rangle = \sum_{j=1}^{k} A_{ij} B_j < 0 \quad \forall i = 1..k.$$

With induction to k:

$k=1$: trivial: choose $B_1 = -A_{11}$ done

Let it hold for $k-1$, then, we can already find $B_1 .. B_{k-1}$ with

$$\sum_{j=1}^{k-1} A_{ij} B_j < 0 \qquad \forall i = 1..k-1.$$

Now, let B_k be such that

$$A_{kk} B_k < -\sum_{j=1}^{k-1} A_{kj} B_j$$

and $A_{kk} B_k < 0$.

Then we see $\sum_{j=1}^{k} A_{ij} B_j < 0 \quad \forall i = 1..k$ $\square$

Note: we need linear independence to apply Gram-Schmidt.
Indeed, the lemma fails to hold for $\{(1,0),(-1,0)\}$, and this set is linearly dependent.

proof of KKT (10.20):

$\subset$ let $v \in T_x E$, with $(x_n) \subset E$, $(\lambda_n) \subset (0,1)$ such that

$$x_n \to x \qquad \text{and} \qquad \frac{x_n - x}{\lambda_n} \to v$$
$$\lambda_n \to 0$$

Then by differentiability of Φ_i,

$$\frac{\Phi_i(x_n) - \Phi_i(x)}{\lambda_n} \longrightarrow \langle \nabla \Phi_i(x), v \rangle$$

for $i = r+1, \dots, k$ we have $\Phi_i(x) = 0$ and $\Phi_i(x_n) \leq 0$ always, therefore we use that the limit preserves the inequality:

$$\frac{\Phi_i(x_n) - \Phi_i(x)}{\lambda_n} = \frac{\Phi_i(x_n)}{\lambda_n} \leq 0$$

so $\langle \nabla \Phi_i(x), v \rangle \leq 0$

This proves $\subset$.

$\supset$ We already can see that if $\langle \nabla \Phi_i(x), v \rangle \leq 0$ for all $i = 1..k$, then by continuity of Φ and openness of

$$\left\{ \tilde{x} \in \mathbb{R}^N : \begin{array}{ll} \Phi_i(\tilde{x}) < \Phi_i(x) & \forall i = 1..r \\ \Phi_i(\tilde{x}) < 0 & \forall i = r+1..k \end{array} \right\}$$

we can pick $N \in \mathbb{N}$ suff. large that
$\Phi_i(x + \frac{1}{n} v)$