Theory of Constrained Optimization. Week 4

Note Title 1/29/2022

Other constraint qualifications

If you feel that the LICQ is natural, I'm going to argue the opposite.

It's easy to break LICQ without changing Ω.

If you do not like the "$x_1^2+x_2^2-1=0 \to (x_1^2+x_2^2-1)^2=0$" example, here is another one:

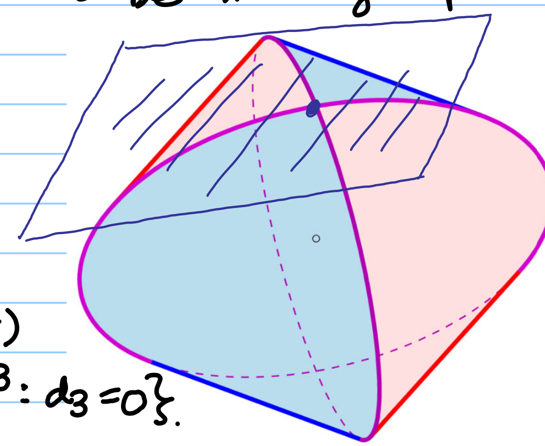
$$\text{write } c_i(x)=0 \quad \text{as} \quad c_i(x)\geq 0 \quad \text{and} \quad \underbrace{-c_i(x)}_{\text{call it } c_{i'}(x)}\leq 0$$

Then $\nabla c_i(x)$ and $\nabla c_{i'}(x) = -\nabla c_i(x)$ are guaranteed to be linearly dependent.

If you still think this is contrived, how about:

HW2

$$\Omega=\left\{x\in\mathbb{R}^3 : \begin{array}{l} x_1^2+x_3^2 \leq 1 \\ x_2^2+x_3^2 \leq 1 \end{array}\right\} \quad \text{at } x^*=\begin{bmatrix}0\\0\\1\end{bmatrix} \; ?$$



In this case, it is easy to argue that $T_\Omega(x^*) = \mathcal{F}(x^*)$
$= \{d\in\mathbb{R}^3 : d_3=0\}.$

One situation in which the linearized feasible direction set $\mathcal{F}(x^*)$ is obviously an adequate representation of the actual feasible set occurs when all the active constraints are already linear, i.e.

$$c_i(x) = a_i^T x + b_i \;, \quad a_i \in \mathbb{R}^n,\; b_i \in \mathbb{R}.$$

Lemma 12.7.

Suppose that at some $x^* \in \Omega$, all active constraints $c_i(\cdot)$, $i \in \mathcal{A}(x^*)$, are linear functions. Then $\mathcal{F}(x^*) = T_\Omega(x^*)$.

We know that $T_\Omega(x^*) \subset \mathcal{F}(x^*)$ is always true.

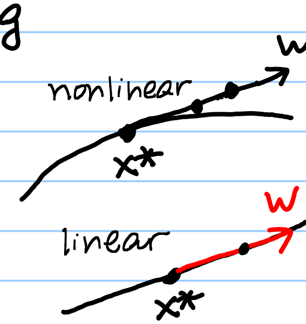
To prove $\mathcal{F}(x^*) \subset T_\Omega(x^*)$ let $w \in \mathcal{F}(x^*)$, we need to find $z_k \in \Omega$, $t_k > 0$ s.t.

$$\frac{z_k - x^*}{t_k} \to w.$$

Recall that, in general (when the constraints are nonlinear), choosing

$$z_k = x^* + \frac{1}{k} w$$

does not always work, but it will work in the linear case.



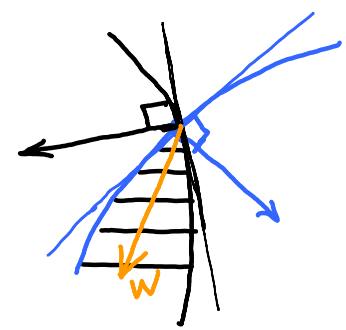
(See NBW Pg 338 if you want to see the details spelt out.)

- The condition that all active constraints be linear is another possible CQ.
- It is neither weaker nor stronger than LICQ.

The following MFCQ condition is weaker than LICQ (why?), and it can guarantee $T_\Omega(x^*) = \mathcal{F}(x^*)$.

Definition 12.6 (MFCQ). (1967 ?)

We say that the Mangasarian–Fromovitz constraint qualification (MFCQ) *holds if there exists a vector* $w \in \mathbb{R}^n$ *such that*



$$\begin{aligned} \nabla c_i(x^*)^T w &> 0, \quad \text{for all } i \in \mathcal{A}(x^*) \cap \mathcal{I}, \\ \nabla c_i(x^*)^T w &= 0, \quad \text{for all } i \in \mathcal{E}, \end{aligned}$$

and the set of equality constraint gradients $\{\nabla c_i(x^*), i \in \mathcal{E}\}$ *is linearly independent.*

When $\mathcal{E} = \emptyset$ and c_i are concave functions (so Ω is convex), then MFCQ is guaranteed by Slater's condition : $\exists\, \hat{x}$ s.t. $c_i(\hat{x}) > 0$, $\forall i$. (Proof : choose $w = \hat{x} - x^*$.)

$x^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

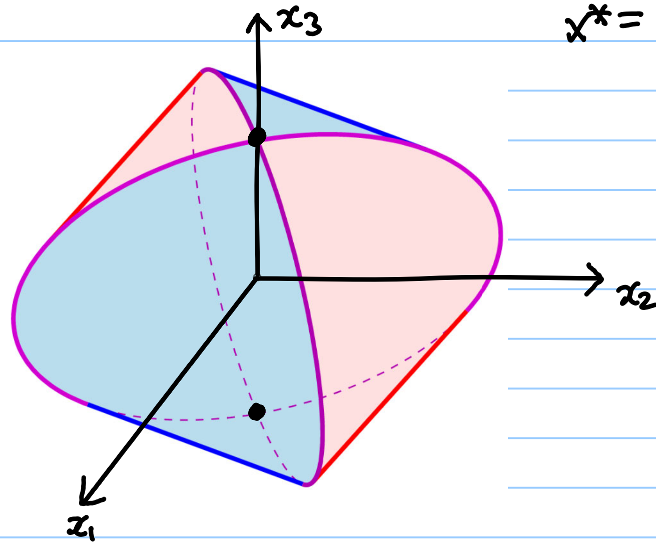
Recall:

both convex

$\Omega := \{ x \in \mathbb{R}^2 : \ x_1^2 - x_2 \geq 0, \ -x_2 \geq 0 \}$ $\qquad \{x^*\} = T_\Omega(x^*) \subsetneq \mathcal{F}(x^*) = x_1\text{-axis}$ $\qquad$ Slater's cond. not satisfied

vs

$\Omega := \{ x \in \mathbb{R}^3 : \ 1 - x_1^2 - x_3^2 \geq 0, \ 1 - x_2^2 - x_3^2 \geq 0 \},$ $\qquad T_\Omega(x^*) = \mathcal{F}(x^*)$ $\qquad$ Slater's cond. satisfied



$x^* = \begin{bmatrix} 0 \\ 0 \\ \pm 1 \end{bmatrix}$

Digression:

Let $p : \mathbb{R}^n \to \mathbb{R}$ be a degree 4 polynomial.

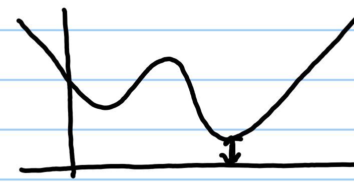
n=1

Is $\min_{x \in \mathbb{R}^n} p(x)$ a convex optimization? Clearly not even in 1-D.

[I gave an example in HW 1 for how minimizing a degree 4 polynomial can show up in imaging science.]

What about:

$$\max_{\gamma \in \mathbb{R}} \gamma \quad \text{s.t.} \quad \underbrace{p(x) - \gamma \geq 0, \ \forall x \in \mathbb{R}^n}_{\text{a very difficult constraint to deal with!}} \ ?$$



a "relaxation"

??

$$\max_{\gamma, Q} \gamma \quad \text{s.t.} \quad \underbrace{p(x) - \gamma = [1, x_1, \dots, x_n, x_1^2, \dots, x_ix_j, \dots, x_n^2] \, Q \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_ix_j \\ \vdots \\ x_n^2 \end{bmatrix}}_{\text{just a set of linear constraints on the entries of } Q}, \quad \underbrace{Q \succeq 0}_{}$$

γ, Q

$\binom{n+2}{2} \times \binom{n+2}{2}$ symmetric matrix

a nonlinear, but convex constraint on the entries of Q!

- The latter optimization problem is an example of a semidefinite program (SDP).
- It is a convex optimization problem: the feasible region

$$\Omega := \left\{ \gamma \in \mathbb{R},\ Q \in \mathbb{R}_{\text{sym}}^{\binom{n+2}{2} \times \binom{n+2}{2}} : \begin{array}{l} \text{a set of linear constraints satisfied} \\ \text{by } \gamma \text{ and } Q \text{, and } Q \succeq 0 \end{array} \right\}$$

viewed as a subset of $\mathbb{R}^{1+1+2+\cdots+\binom{n+2}{2}}$ is convex.

The objective is linear.

- But the constraint "$Q \succeq 0$" isn't quite in the form of $c_i(Q) \geq 0$, $i=1,\dots,m$, (at least not directly.)
- When $n=1$, the "$\approx$" is actually a "$=$", meaning that the SDP problem is equivalent to the original problem of minimizing a deg. 4 polynomial.
- When $n>1$, the relaxed SDP problem is not equivalent to the original problem, but there is a sequence of SDP problems, called Lasserre's Hierachy, of which the solutions converge to that of the original problem (under suitable assumptions.)

SIAM REVIEW
Vol. 49, No. 4, pp. 651–669

A Sum of Squares Approximation of Nonnegative Polynomials*

Jean B. Lasserre†

Abstract. We show that every real nonnegative polynomial f can be approximated as closely as desired (in the l_1-norm of its coefficient vector) by a sequence of polynomials $\{f_\epsilon\}$ that are sums of squares. The novelty is that each f_ϵ has a simple and explicit form in terms of f and ϵ.

Key words. real algebraic geometry, positive polynomials, sum of squares, semidefinite programming

AMS subject classifications. 12E05, 12Y05, 90C22

DOI. 10.1137/070693709

https://www.quantamagazine.org/a-classical-math-problem-gets-pulled-into-the-modern-world-20180523/

Convex Optimization

Stephen Boyd

Department of Electrical Engineering
Stanford University

Lieven Vandenberghe

Electrical Engineering Department
University of California, Los Angeles

CAMBRIDGE
UNIVERSITY PRESS

Back to the basic theory : A geometric viewpoint

Since different constraint functions may define the same Ω, and sometimes Ω is not directly defined by constraint functions, it would be good to phrase the necessity condition purely in terms of (the geometry of) Ω.

Recall the basic argument leading to the KKT conditions :

a cone

with CQ

$$\mathcal{F}(x^*) = \{d : \nabla c_i(x^*)^T d \; = 0,\ i \in \mathcal{E};\ \geq 0,\ i \in \mathcal{I} \cap \mathcal{A}(x^*)\}$$

x^* solves $\min_x f(x)$

s.t. $c_i(x) = 0,\ i \in \mathcal{E}$

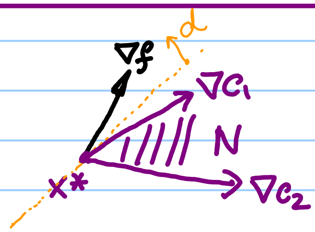
$c_i(x) \geq 0\ \ i \in \mathcal{I}$

$$\Rightarrow \nexists\, d \in \mathbb{R}^n \text{ s.t. } d \in T_\Omega(x^*) \text{ and } \nabla f(x^*)^T d < 0$$

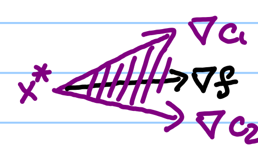
$\Updownarrow$ Farkas' lemma (when $T_\Omega(x^*) = \mathcal{F}(x^*)$)

$$\nabla f(x^*) \in N = \{\sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla c_i(x^*) : \lambda_i \geq 0 \text{ for } i \in \mathcal{A}(x^*) \cap \mathcal{I}\}$$

also a cone



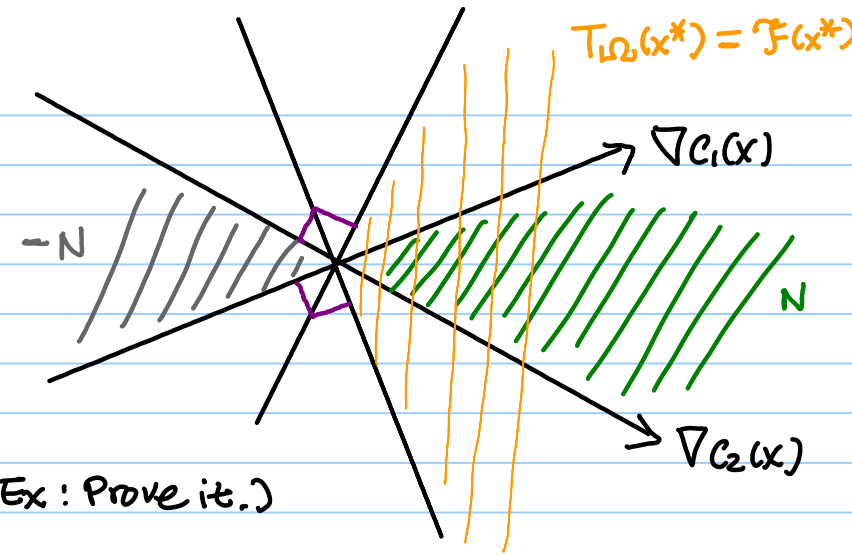
$\Rightarrow$ x^* cannot be a local minimizer



(there is a chance)

A relation between N and $\mathcal{F}(x^*)$:

$$\boxed{\begin{aligned} -N &= \{ v \in \mathbb{R}^n : v^T w \le 0,\ w \in \mathcal{F}(x^*) \} \\ &= \text{the set of all vectors that make an angle of at least } \tfrac{\pi}{2} \text{ with every tangent vector} \end{aligned}}$$



(Ex: Prove it.)

Definition 12.7.

The normal cone *to the set* Ω *at the point* $x \in \Omega$ *is defined as*

$$N_\Omega(x) = \{v \mid v^T w \le 0 \text{ for all } w \in T_\Omega(x)\}, \tag{12.77}$$

where $T_\Omega(x)$ *is the tangent cone of Definition 12.2. Each vector* $v \in N_\Omega(x)$ *is said to be a* normal vector.

With only equality constraints and assuming LICQ, Ω is a $n-|\mathcal{E}|$ dim. smooth surface near x^*, $T_\Omega(x^*)$ is the usual tangent plane of Ω at x^*, $N_\Omega(x^*) = T_\Omega(x^*)^\perp$.

Theorem 12.8.

Suppose that x^* is a local minimizer of f in Ω. Then

$$-\nabla f(x^*) \in N_\Omega(x^*). \tag{12.78}$$

Proof: Given any $d \in T_\Omega(x^*)$, $\exists\, t_k > 0$, $z_k \in \Omega$ s.t. $z_k = x^* + t_k d + o(t_k)$.

If x^* is a local minimizer,

$$f(z_k) \geq f(x^*).$$

Since f is C^1,

$$\begin{aligned} f(z_k) - f(x^*) &= \nabla f(x^*)^T (z_k - x^*) + o(\|z_k - x^*\|) \\ &= t_k \nabla f(x^*)^T d + o(t_k) \end{aligned}$$

(we used it before)

So $\underbrace{\dfrac{f(z_k) - f(x^*)}{t_k}}_{\geq 0} = \nabla f(x^*)^T d + \underbrace{\dfrac{o(t_k)}{t_k}}_{\to 0 \text{ as } k \to \infty} \implies \nabla f(x^*)^T d \geq 0$

So $-\nabla f(x^*)^T d \leq 0 \quad \forall d \in T_\Omega(x^*)$ ie $-\nabla f(x^*) \in N_\Omega(x^*)$.

Q.E.D.

The result above is purely based on the geometry of Ω. If now Ω is defined by constraint functions in the form we have been assuming, then the following holds (as expected):

Lemma 12.9.

Suppose that the LICQ assumption (Definition 12.4) holds at x^*. Then t the normal cone $N_\Omega(x^*)$ is simply $-N$, where N is the set defined in (12.50).

PROOF. The proof follows from Farkas' Lemma (Lemma 12.4) and Definition 12.7 of $N_\Omega(x^*)$. From Lemma 12.4, we have that

$$g \in N \quad \Rightarrow \quad g^T d \geq 0 \ \text{ for all } d \in \mathcal{F}(x^*).$$

Since we have $\mathcal{F}(x^*) = T_\Omega(x^*)$ from Lemma 12.2, it follows by switching the sign of this expression that

$$g \in -N \quad \Rightarrow \quad g^T d \leq 0 \ \text{ for all } d \in T_\Omega(x^*).$$

We conclude from Definition 12.7 that $N_\Omega(x^*) = -N$, as claimed. □

Lagrange multipliers and sensitivity

Since $\nabla f(x^*) = \sum_i \lambda_i^* \nabla c_i(x^*)$, it's not surprising that λ_i^* has something to do with how much $f(x^*)$ changes when $c_i(x^*)$ changes.

For instance, if $i \notin \mathcal{A}(x^*)$, $\lambda_i^* = 0$, which is consistent with that $f(x^*)$ is insensitive to small changes in $c_i(x^*)$.

If $i \in \mathcal{A}(x^*) \cap \mathcal{I}$, assume that the constraint $c_i(x) \geq 0$ is changed to

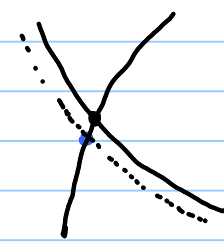
$$c_i(x) \geq \varepsilon,$$

call the perturbed solution

$$x^*(\varepsilon)$$

then under mild assumptions

$$\frac{df(x^*(\varepsilon))}{d\varepsilon} = \lambda_i.$$



Duality

It begins with the following trick for obtaining lower bounds of

$$f^* = \min_x f(x) \text{ s.t. } \begin{array}{ll} c_i(x) = 0 & i \in \mathcal{E} \\ c_i(x) \geq 0 & i \in \mathcal{I}. \end{array}$$

If λ is st $\lambda_i \geq 0$ for $i \in \mathcal{I}$ (no sign constraint for λ_i, $i \in \mathcal{E}$)
then

$$\underbrace{f(x) - \sum_i \lambda_i c_i(x)}_{\mathcal{L}(x,\lambda)} \leq f(x) \text{ when } \begin{array}{ll} c_i(x) = 0 & i \in \mathcal{E} \\ c_i(x) \geq 0 & i \in \mathcal{I} \end{array}$$

So if $g(\lambda) := \inf_x \mathcal{L}(x,\lambda)$ s.t. $\begin{array}{ll} c_i(x) = 0 & i \in \mathcal{E} \\ c_i(x) \geq 0 & i \in \mathcal{I}. \end{array}$, then $g(\lambda) \leq f^*$.

Now "throw the constraints away" and consider $q(\lambda) := \inf_x \mathcal{L}(x,\lambda)$.
Obviously

$$q(\lambda) \leq g(\lambda) \leq f^*. \qquad \forall \lambda \text{ st } \lambda_i \geq 0, i \in \mathcal{I}.$$

called the dual problem

We can then tighten the lower bound by taking $q^* := \boxed{\max_{\substack{\lambda \\ \lambda_i \geq 0, i \in \mathcal{I}}} q(\lambda)}$
and we still have

$$q^* \leq f^*. \qquad (f^* - q^* = \text{the duality gap})$$

The dual problem, being a "max-min" problem, looks pretty nasty / useless. But then it has a very nice feature:

Thm: $q : \mathbb{R}^{\overset{=|\mathcal{E}\cup\mathcal{I}|}{m}} \to \mathbb{R}\cup\{-\infty\}$ is concave. The (natural) domain of q,
$\mathcal{D} := \{\lambda \in \mathbb{R}^m : q(\lambda) > -\infty\}$, is convex.

Proof: Let $\lambda^0, \lambda^1 \in \mathbb{R}^m$, $x \in \mathbb{R}^n$, $\alpha \in [0,1]$.

We have $\mathcal{L}(x, (1-\alpha)\lambda^0 + \alpha\lambda^1) = (1-\alpha)\mathcal{L}(x,\lambda^0) + \alpha\mathcal{L}(x,\lambda^1)$.

$$
\begin{aligned}
q((1-\alpha)\lambda^0 + \alpha\lambda^1) &= \inf_x \mathcal{L}(x, (1-\alpha)\lambda^0 + \alpha\lambda^1) \\
&= \inf_x (1-\alpha)\mathcal{L}(x,\lambda^0) + \alpha\mathcal{L}(x,\lambda^1) \\
&\overset{\text{why?}}{\geq} (1-\alpha)\inf_x \mathcal{L}(x,\lambda^0) + \alpha \inf_x \mathcal{L}(x,\lambda^1) = (1-\alpha)q(\lambda^0) + \alpha q(\lambda^1).
\end{aligned}
$$

If $\lambda^0, \lambda^1 \in \mathcal{D}$, $q(\lambda^0) > -\infty$, $q(\lambda^1) > -\infty$, so the concavity of q we just established also shows $q((1-\alpha)\lambda^0 + \alpha\lambda^1) > -\infty$, i.e. $(1-\alpha)\lambda^0 + \alpha\lambda^1 \in \mathcal{D}$.

Q.E.D.

Note: Whether the original problem is convex or not is irrelevant to this result.

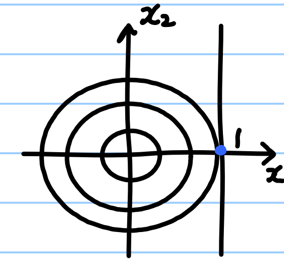
both convex

$$\max_{\lambda} q(\lambda),\ \lambda \in \mathcal{D} \cap \{\lambda \in \mathbb{R}^m : \lambda_i \geq 0,\ i \in \mathcal{I}\}$$

($q^* = -\infty$ if the intersection is empty)

The fact that $q^* \leq f^*$ is called the **weak duality** property/theorem.

The value $f^* - q^*$ is called the **duality gap**.



E.g. $\displaystyle\min_{x \in \mathbb{R}^2} \tfrac{1}{2}(x_1^2 + x_2^2)$ s.t. $x_1 - 1 \geq 0$ Solution at $x^* = \begin{bmatrix}1\\0\end{bmatrix}$, $f^* = \frac{1}{2}$

$$\nabla f(x^*) = \begin{bmatrix}1\\0\end{bmatrix} = \underset{\lambda^*=}{1}\begin{bmatrix}1\\0\end{bmatrix}$$

$$\mathcal{L}(x,\lambda) = \tfrac{1}{2}(x_1^2 + x_2^2) - \lambda_1(x_1 - 1)$$ ← convex for any fixed λ_1

min. when $\nabla_x \mathcal{L}(x,\lambda) = 0$, i.e. $x_1 - \lambda_1 = 0$, $x_2 = 0$

$$q(\lambda_1) = \min_x \mathcal{L}(x,\lambda_1) = \mathcal{L}\left(\begin{bmatrix}\lambda_1\\0\end{bmatrix}, \lambda_1\right) = -\tfrac{1}{2}\lambda_1^2 + \lambda_1$$ ← concave

$q' = -\lambda_1 + 1$
$= 0$ when $\lambda_1 = 1$

Dual problem: $\displaystyle\max_{\lambda_1 \geq 0} -\tfrac{1}{2}\lambda_1^2 + \lambda_1$. Solution at $\lambda_1 = 1$.

$q^* = \frac{1}{2} = f^*$, duality gap $= 0$.

Eg (HW) min $x_1^2 - 3x_2^2$ s.t. $x_1 = x_2^3$. ← neither f nor Ω is convex

$q^* = -\infty$, $f^* = -2$.

Theorem (No duality gap for convex problems)

Suppose that $\bar{x}$ is a solution of $\underbrace{\min f(x) \text{ st } c_i(x) \geq 0}_{\text{a convex problem}}$, $f, -c_i$ are convex and differentiable at $\bar{x}$.

Then any $\bar{\lambda}$ for which $(\bar{x}, \bar{\lambda})$ satisfies the KKT conditions is a solution of the dual problem, and $q(\bar{\lambda}) = f(\bar{x})$.

$c(x) = \begin{bmatrix} c_1(x) \\ \vdots \\ c_m(x) \end{bmatrix}$, $\nabla c(x) = [\nabla c_1(x), \cdots, \nabla c_m(x)]$

Proof: Suppose

(KKT) $\nabla f(\bar{x}) - \nabla c(\bar{x})^T \bar{\lambda} = 0$, $c(\bar{x}) \geq 0$, $\bar{\lambda} \geq 0$, $\bar{\lambda}_i c_i(\bar{x}) = 0$, $i = 1, \cdots, m$.

$\mathcal{L}(\cdot, \bar{\lambda})$ is convex and diff. at $\bar{x}$, so, for any $x \in \mathbb{R}^n$,

$$\mathcal{L}(x, \bar{\lambda}) \geq \mathcal{L}(\bar{x}, \bar{\lambda}) + \underbrace{\nabla \mathcal{L}(\bar{x}, \bar{\lambda})^T}_{=0} (x - \bar{x}) \overset{\text{KKT}}{=} \mathcal{L}(\bar{x}, \bar{\lambda}).$$

$$q(\bar{\lambda}) = \inf_x \mathcal{L}(x, \bar{\lambda}) = \mathcal{L}(\bar{x}, \bar{\lambda}) = f(\bar{x}) - \underbrace{\bar{\lambda}^T c(\bar{x})}_{=0} \overset{\text{KKT}}{=} f(\bar{x}).$$

By weak duality, this also means $\bar{\lambda}$ is a maximizer of q, i.e. a sol. of the dual problem.

Q.E.D.

(A partial converse)

Suppose that $f, -c_i : \mathbb{R}^n \to \mathbb{R}$ are convex C^1 functions,

$\bar{x}$ solves $\overbrace{\min_x f(x) \text{ st } c_i(x) \geq 0,\ i=1,\cdots,m}^{(*)}$, and LICQ holds at $\bar{x}$.

Suppose that $\hat{\lambda}$ solves the dual problem, and $\inf_x \mathcal{L}(\cdot,\hat{\lambda}) = \mathcal{L}(\hat{x},\hat{\lambda})$.

If $\mathcal{L}(\cdot,\hat{\lambda})$ is (not only convex but) strictly convex, then $\hat{x} = \bar{x}$ and $f(\bar{x}) = \mathcal{L}(\hat{x},\hat{\lambda})$.

Proof: By assumptions, $\exists\, \bar{\lambda}$ st. $(\bar{x},\bar{\lambda})$ satisfies the KKT conditions.

By the previous theorem and its proof, $\bar{\lambda}$ solves the dual problem also, so that

$$\mathcal{L}(\bar{x},\bar{\lambda}) = q(\bar{\lambda}) = q(\hat{\lambda}) = \mathcal{L}(\hat{x},\hat{\lambda}).$$

Since $\hat{x} = \operatorname{argmin}_x \mathcal{L}(x,\hat{\lambda})$, $\nabla_x \mathcal{L}(\hat{x},\hat{\lambda}) = 0$. Now, assume that $\bar{x} \neq \hat{x}$. By strict convexity,

$$\mathcal{L}(\bar{x},\hat{\lambda}) - \mathcal{L}(\hat{x},\hat{\lambda}) > \nabla_x \mathcal{L}(\hat{x},\hat{\lambda})^T(\bar{x}-\hat{x}) = 0.$$

So

$$\mathcal{L}(\bar{x},\hat{\lambda}) > \mathcal{L}(\hat{x},\hat{\lambda}) = \mathcal{L}(\bar{x},\bar{\lambda}),$$

which means

$$-\hat{\lambda}^T c(\bar{x}) > -\bar{\lambda}^T c(\bar{x}) \overset{\text{KKT}}{=} 0. \quad \text{But } \hat{\lambda} \geq 0,\ c(\bar{x}) \geq 0,\ \Rightarrow\!\Leftarrow.$$

Q.E.D.

Note: under the setting of this theorem, $\hat{x}$ is the unique minimizer of $\mathcal{L}(\cdot,\hat{\lambda})$, and hence $\bar{x} = \hat{x}$ is the unique solution of $(*)$.

Wolfe dual

$$\max_{\lambda} \inf_{x} \overbrace{f(x) - \sum \lambda_i c_i(x)}^{= \mathcal{L}(x,\lambda) \text{ convex in } x} \quad \text{s.t. } \lambda \geq 0$$

can be recast as

$$\boxed{\max_{x,\lambda} \mathcal{L}(x,\lambda) \quad \text{s.t. } \underbrace{\nabla_x \mathcal{L}(x,\lambda) = 0}_{\text{guarantees that } \mathcal{L}(\cdot,\lambda) \text{ is minimized}}, \ \lambda \geq 0} \leftarrow \text{Wolfe dual}$$

Example (Linear Programming)

$$\min c^T x \quad \text{st} \quad Ax - b \geq 0$$

$$q(\lambda) = \inf_x [c^T x - \lambda^T (Ax - b)] = \inf_x [(c - A^T\lambda)x + b^T\lambda] = \begin{cases} -\infty & \text{if } A^T\lambda \neq c \\ b^T\lambda & \text{if } A^T\lambda = c \end{cases}$$

In maximizing q, we can exclude λ for which $A^T\lambda \neq c$ from consideration (the max obviously cannot be attained at a point λ for which $q(\lambda) = -\infty$.) So, we may write the dual problem as:

$$\max b^T\lambda \quad \text{s.t.} \quad A^T\lambda = c, \ \lambda \geq 0.$$ ← also a LP!

Ex: Explain: if the primal LP is $\begin{array}{l} \min c^T x \\ \text{st. } Ax \geq b \\ \quad x \geq 0 \end{array}$, then the dual LP is $\begin{array}{l} \max b^T\lambda \\ \text{s.t. } A^T\lambda \leq c \\ \quad \lambda \geq 0. \end{array}$

What if you take the dual of the dual?

Wolfe dual: $\max_{x,\lambda} \underbrace{c^T x - \lambda^T (Ax - b)}_{= (c - A^T\lambda)x + b^T\lambda}$ st $A^T\lambda = c$, $\lambda \geq 0$. By sub. the constraint $A^T\lambda - c = 0$ into the objective, we get back the same dual.

Example (convex quadratic program)

$$\min_x \tfrac{1}{2}x^TGx + c^Tx \quad \text{s.t. } Ax-b \geq 0 \qquad (G \succ 0)$$

$$q(\lambda) = \inf_x \underbrace{\tfrac{1}{2}x^TGx + c^Tx - \lambda^T(Ax-b)}_{\mathcal{L}(x,\lambda)\text{, strictly convex in } x,\ \forall\lambda}$$

$\mathcal{L}(\cdot,\lambda)$ is minimized at the $x \in \mathbb{R}^n$ that satisfies $\underbrace{\nabla_x \mathcal{L}(x,\lambda)}_{= Gx + c - A^T\lambda} = 0$, or when

$$x = G^{-1}(A^T\lambda - c).$$

So,

$$q(\lambda) = \tfrac{1}{2}x^TGx + \underbrace{(c - A^T\lambda)^T x}_{-(Gx)^Tx = -x^TGx} + b^T\lambda = -\tfrac{1}{2}x^TGx + b^T\lambda \Big|_{x = G^{-1}(A^T\lambda - c)}$$

$$= -\tfrac{1}{2}(A^T\lambda - c)^T G^{-1}(A^T\lambda - c) + b^T\lambda.$$

The dual problem is $\max_{\lambda \geq 0} -\frac{1}{2}(A^T\lambda - c)^T G^{-1}(A^T\lambda - c) + b^T\lambda.$

concave in x
linear in λ

Wolfe dual: $\max_{\lambda, x} \frac{1}{2}x^TGx + c^Tx - \lambda^T(Ax-b)$ s.t. $Gx + c - A^T\lambda = 0$, $\lambda \geq 0$, equiv to $\max_{\lambda,x} -\frac{1}{2}x^TGx + \lambda^Tb$ s.t. $Gx + c - A^T\lambda = 0$, $\lambda \geq 0$.

(Note: The Wolfe dual is a well-defined convex problem as long as $G \succeq 0$.)