MATH4230 - Optimization Theory - 2019/20

Tieyong Zeng
zeng@math.cuhk.edu.hk
March, 2020

Plan (March 10-11, 2020)

1. Review of subgradient
2. Duality
3. Kuhn-Tucker theorem

Subgradients

Ryan Tibshirani
Convex Optimization 10-725

Last time: gradient descent

Consider the problem

$$\min_x \ f(x)$$

for f convex and differentiable, $\mathrm{dom}(f) = \mathbb{R}^n$. Gradient descent: choose initial $x^{(0)} \in \mathbb{R}^n$, repeat

$$x^{(k)} = x^{(k-1)} - t_k \cdot \nabla f(x^{(k-1)}), \quad k = 1, 2, 3, \dots$$

Step sizes t_k chosen to be fixed and small, or by backtracking line search

If ∇f is Lipschitz, gradient descent has convergence rate $O(1/\epsilon)$. Downsides:

- Requires f differentiable
- Can be slow to converge

Outline

Today: crucial mathematical underpinnings!

- Subgradients
- Examples
- Properties
- Optimality characterizations

Subgradients

Recall that for convex and differentiable f,

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \text{for all } x, y$$

That is, linear approximation always underestimates f

A subgradient of a convex function f at x is any $g \in \mathbb{R}^n$ such that

$$f(y) \geq f(x) + g^T (y - x) \quad \text{for all } y$$

- Always exists[1]
- If f differentiable at x, then $g = \nabla f(x)$ uniquely
- Same definition works for nonconvex f (however, subgradients need not exist)

[1] On the relative interior of $\mathrm{dom}(f)$

Example 2.38 Let $p(x) := \|x\|$ be the Euclidean norm function on $\mathbb{R}^n$. Then we have

$$\partial p(x) = \begin{cases} I\!B & \text{if } x = 0, \\ \left\{ \dfrac{x}{\|x\|} \right\} & \text{otherwise.} \end{cases}$$

To verify this, observe first that the Euclidean norm function p is differentiable at any nonzero point with $\nabla p(x) = \dfrac{x}{\|x\|}$ as $x \neq 0$. It remains to calculate its subdifferential at $x = 0$. To proceed by definition (2.13), we have that $v \in \partial p(0)$ if and only if

$$\langle v, x \rangle = \langle v, x - 0 \rangle \leq p(x) - p(0) = \|x\| \text{ for all } x \in \mathbb{R}^n.$$

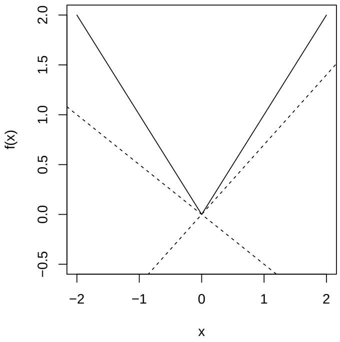
Letting $x = v$ gives us $\langle v, v \rangle \leq \|v\|$, which implies that $\|v\| \leq 1$, i.e., $v \in I\!B$. Now take $v \in I\!B$ and deduce from the Cauchy-Schwarz inequality that

$$\langle v, x - 0 \rangle = \langle v, x \rangle \leq \|v\| \cdot \|x\| \leq \|x\| = p(x) - p(0) \text{ for all } x \in \mathbb{R}^n$$

and thus $v \in \partial p(0)$, which shows that $\partial p(0) = I\!B$.

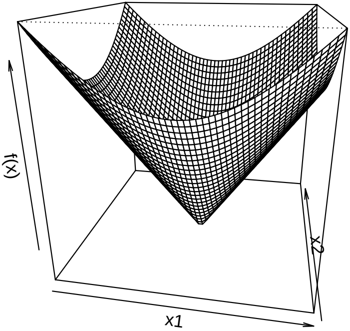
Examples of subgradients

Consider $f : \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$



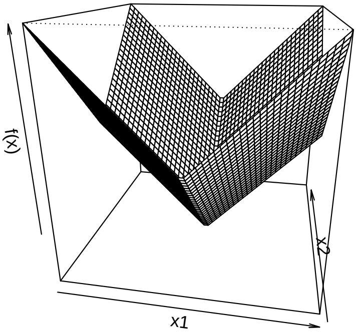
- For $x \neq 0$, unique subgradient $g = \mathrm{sign}(x)$
- For $x = 0$, subgradient g is any element of $[-1, 1]$

Consider $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \|x\|_2$



- For $x \neq 0$, unique subgradient $g = x/\|x\|_2$
- For $x = 0$, subgradient g is any element of $\{z : \|z\|_2 \leq 1\}$

Consider $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \|x\|_1$



- For $x_i \neq 0$, unique ith component $g_i = \mathrm{sign}(x_i)$
- For $x_i = 0$, ith component g_i is any element of $[-1, 1]$

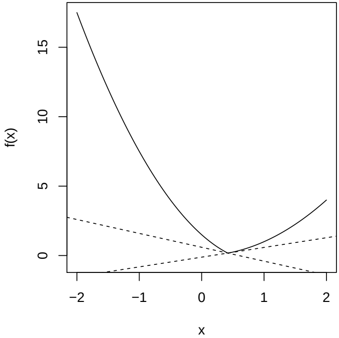
Theorem 2.17 (Dubovitskii-Milyutin) *Let* $f_1, \cdots, f_m : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions and let* x_0 *be a point in* $\cap_{i=1}^m \mathrm{int\ dom}\ f_i$. *Let* $f : \mathbb{R}^n \to (-\infty, +\infty]$ *be given by*

$$f(x) := \max_{1 \le i \le m} f_i(x)$$

and let $I(x_0)$ *be the (nonempty) set of all* $i \in \{1, \cdots, m\}$ *for which* $f_i(x_0) = f(x_0)$. *Then*

$$\partial f(x_0) = \mathrm{co}\ \cup_{i \in I(x_0)} \partial f_i(x_0).$$

Consider $f(x) = \max\{f_1(x), f_2(x)\}$, for $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$ convex, differentiable



- For $f_1(x) > f_2(x)$, unique subgradient $g = \nabla f_1(x)$
- For $f_2(x) > f_1(x)$, unique subgradient $g = \nabla f_2(x)$
- For $f_1(x) = f_2(x)$, subgradient g is any point on line segment between $\nabla f_1(x)$ and $\nabla f_2(x)$

Subdifferential

Set of all subgradients of convex f is called the subdifferential:

$$\partial f(x) = \{g \in \mathbb{R}^n : g \text{ is a subgradient of } f \text{ at } x\}$$

- Nonempty (only for convex f)
- $\partial f(x)$ is closed and convex (even for nonconvex f)
- If f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$
- If $\partial f(x) = \{g\}$, then f is differentiable at x and $\nabla f(x) = g$

Connection to convex geometry

Convex set $C \subseteq \mathbb{R}^n$, consider indicator function $I_C : \mathbb{R}^n \to \mathbb{R}$,

$$I_C(x) = I\{x \in C\} = \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{if } x \notin C \end{cases}$$

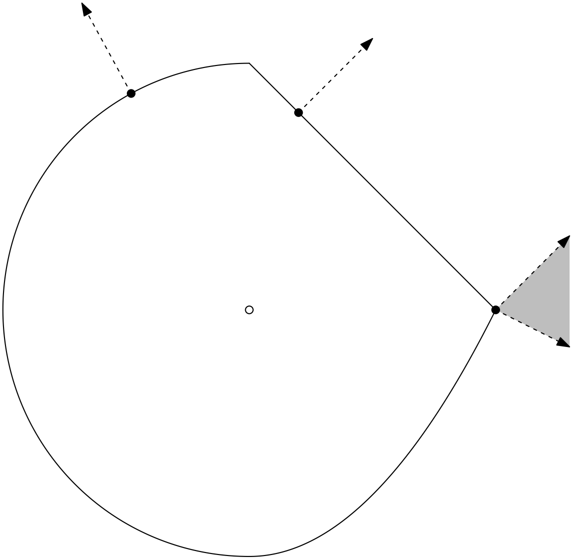
For $x \in C$, $\partial I_C(x) = \mathcal{N}_C(x)$, the normal cone of C at x is, recall

$$\mathcal{N}_C(x) = \{g \in \mathbb{R}^n : g^T x \geq g^T y \text{ for any } y \in C\}$$

Why? By definition of subgradient g,

$$I_C(y) \geq I_C(x) + g^T(y - x) \quad \text{for all } y$$

- For $y \notin C$, $I_C(y) = \infty$
- For $y \in C$, this means $0 \geq g^T(y - x)$



Subgradient calculus

Basic rules for convex functions:

- Scaling: $\partial(af) = a \cdot \partial f$ provided $a > 0$
- Addition: $\partial(f_1 + f_2) = \partial f_1 + \partial f_2$
- Affine composition: if $g(x) = f(Ax + b)$, then

$$\partial g(x) = A^T \partial f(Ax + b)$$

- Finite pointwise maximum: if $f(x) = \max_{i=1,\ldots,m} f_i(x)$, then

$$\partial f(x) = \text{conv}\left(\bigcup_{i: f_i(x) = f(x)} \partial f_i(x) \right)$$

convex hull of union of subdifferentials of active functions at x

https://doi.org/10.1080/02331934.2015.1105225

- General composition: if

$$f(x) = h\big(g(x)\big) = h\big(g_1(x), \ldots, g_k(x)\big)$$

where $g : \mathbb{R}^n \to \mathbb{R}^k$, $h : \mathbb{R}^k \to \mathbb{R}$, $f : \mathbb{R}^n \to \mathbb{R}$, h is convex and nondecreasing in each argument, g is convex, then

$$\partial f(x) \supseteq \Big\{ p_1 q_1 + \cdots + p_k q_k : \\ p \in \partial h(g(x)),\ q_i \in \partial g_i(x),\ i = 1, \ldots, k \Big\}$$

- General pointwise maximum: if $f(x) = \max_{s \in S} f_s(x)$, then

$$\partial f(x) \supseteq \mathrm{cl} \left\{ \mathrm{conv} \left(\bigcup_{s : f_s(x) = f(x)} \partial f_s(x) \right) \right\}$$

Under some regularity conditions (on S, f_s), we get equality

- Norms: important special case. To each norm $\|\cdot\|$, there is a dual norm $\|\cdot\|_*$ such that

$$\|x\| = \max_{\|z\|_* \le 1} z^T x$$

(For example, $\|\cdot\|_p$ and $\|\cdot\|_q$ are dual when $1/p + 1/q = 1$.) In fact, for $f(x) = \|x\|$ (and $f_z(x) = z^T x$), we get equality:

$$\partial f(x) = \mathrm{cl}\left\{\mathrm{conv}\left(\bigcup_{z: f_z(x) = f(x)} \partial f_z(x)\right)\right\}$$

Note that $\partial f_z(x) = z$. And if z_1, z_2 each achieve the max at x, which means that $z_1^T x = z_2^T x = \|x\|$, then by linearity, so will $tz_1 + (1-t)z_2$ for any $t \in [0,1]$. Thus

$$\partial f(x) = \operatorname*{argmax}_{\|z\|_* \le 1} z^T x$$

Optimality condition

For any f (convex or not),

$$f(x^\star) = \min_x \; f(x) \quad \Longleftrightarrow \quad 0 \in \partial f(x^\star)$$

That is, $x^\star$ is a minimizer if and only if 0 is a subgradient of f at $x^\star$. This is called the subgradient optimality condition

Why? Easy: $g = 0$ being a subgradient means that for all y

$$f(y) \geq f(x^\star) + 0^T(y - x^\star) = f(x^\star)$$

Note the implication for a convex and differentiable function f, with $\partial f(x) = \{\nabla f(x)\}$

Derivation of first-order optimality

Example of the power of subgradients: we can use what we have learned so far to derive the first-order optimality condition. Recall

$$\min_x \; f(x) \;\; \text{subject to} \;\; x \in C$$

is solved at x, for f convex and differentiable, if and only if

$$\nabla f(x)^T (y - x) \geq 0 \;\; \text{for all} \;\; y \in C$$

Intuitively: says that gradient increases as we move away from x. How to prove it? First recast problem as

$$\min_x \; f(x) + I_C(x)$$

Now apply subgradient optimality: $0 \in \partial(f(x) + I_C(x))$

Observe

$$
\begin{aligned}
0 \in \partial\big(f(x) + I_C(x)\big) & \\
&\iff 0 \in \{\nabla f(x)\} + \mathcal{N}_C(x) \\
&\iff -\nabla f(x) \in \mathcal{N}_C(x) \\
&\iff -\nabla f(x)^T x \geq -\nabla f(x)^T y \text{ for all } y \in C \\
&\iff \nabla f(x)^T (y - x) \geq 0 \text{ for all } y \in C
\end{aligned}
$$

as desired

Note: the condition $0 \in \partial f(x) + \mathcal{N}_C(x)$ is a fully general condition for optimality in convex problems. But it's not always easy to work with (KKT conditions, later, are easier)

Example: lasso optimality conditions

Given $y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n \times p}$, lasso problem can be parametrized as

$$\min_{\beta} \ \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1$$

where $\lambda \geq 0$. Subgradient optimality:

$$\begin{aligned} 0 \in \partial\Big(\frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1\Big) \\ &\iff \ 0 \in -X^T(y - X\beta) + \lambda\partial\|\beta\|_1 \\ &\iff \ X^T(y - X\beta) = \lambda v \end{aligned}$$

for some $v \in \partial\|\beta\|_1$, i.e.,

$$v_i \in \begin{cases} \{1\} & \text{if } \beta_i > 0 \\ \{-1\} & \text{if } \beta_i < 0 \\ [-1, 1] & \text{if } \beta_i = 0 \end{cases}, \quad i = 1, \ldots, p$$

Write $X_1, \ldots, X_p$ for columns of X. Then our condition reads:

$$\begin{cases} X_i^T(y - X\beta) = \lambda \cdot \text{sign}(\beta_i) & \text{if } \beta_i \neq 0 \\ |X_i^T(y - X\beta)| \leq \lambda & \text{if } \beta_i = 0 \end{cases}$$

Note: subgradient optimality conditions don't lead to closed-form expression for a lasso solution ... however they do provide a way to check lasso optimality

They are also helpful in understanding the lasso estimator; e.g., if $|X_i^T(y - X\beta)| < \lambda$, then $\beta_i = 0$ (used by screening rules, later?)

Example: soft-thresholding

Simplfied lasso problem with $X = I$:

$$\min_{\beta} \ \frac{1}{2}\|y-\beta\|_2^2 + \lambda\|\beta\|_1$$

This we can solve directly using subgradient optimality. Solution is $\beta = S_\lambda(y)$, where S_λ is the soft-thresholding operator:

$$[S_\lambda(y)]_i = \begin{cases} y_i - \lambda & \text{if } y_i > \lambda \\ 0 & \text{if } -\lambda \le y_i \le \lambda \\ y_i + \lambda & \text{if } y_i < -\lambda \end{cases}, \quad i = 1, \ldots, n$$

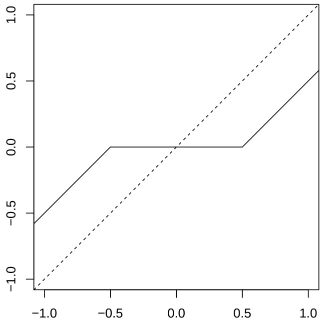
Check: from last slide, subgradient optimality conditions are

$$\begin{cases} y_i - \beta_i = \lambda \cdot \mathrm{sign}(\beta_i) & \text{if } \beta_i \neq 0 \\ |y_i - \beta_i| \le \lambda & \text{if } \beta_i = 0 \end{cases}$$

Now plug in $\beta = S_\lambda(y)$ and check these are satisfied:

- When $y_i > \lambda$, $\beta_i = y_i - \lambda > 0$, so $y_i - \beta_i = \lambda = \lambda \cdot 1$
- When $y_i < -\lambda$, argument is similar
- When $|y_i| \leq \lambda$, $\beta_i = 0$, and $|y_i - \beta_i| = |y_i| \leq \lambda$

Soft-thresholding in one variable:



Example: distance to a convex set

Recall the distance function to a closed, convex set C:

$$\mathrm{dist}(x, C) = \min_{y \in C} \|y - x\|_2$$

This is a convex function. What are its subgradients?

Write $\mathrm{dist}(x, C) = \|x - P_C(x)\|_2$, where $P_C(x)$ is the projection of x onto C. It turns out that when $\mathrm{dist}(x, C) > 0$,

$$\partial \mathrm{dist}(x, C) = \left\{ \frac{x - P_C(x)}{\|x - P_C(x)\|_2} \right\}$$

Only has one element, so in fact $\mathrm{dist}(x, C)$ is differentiable and this is its gradient

We will only show one direction, i.e., that

$$\frac{x - P_C(x)}{\|x - P_C(x)\|_2} \in \partial \mathrm{dist}(x, C)$$

Write $u = P_C(x)$. Then by first-order optimality conditions for a projection,

$$(x-u)^T(y-u) \le 0 \quad \text{for all } y \in C$$

Hence

$$C \subseteq H = \{y : (x-u)^T(y-u) \le 0\}$$

Claim:

$$\mathrm{dist}(y, C) \ge \frac{(x-u)^T(y-u)}{\|x-u\|_2} \quad \text{for all } y$$

Check: first, for $y \in H$, the right-hand side is ≤ 0

Now for $y \notin H$, we have $(x-u)^T(y-u) = \|x-u\|_2\|y-u\|_2 \cos\theta$ where θ is the angle between $x-u$ and $y-u$. Thus

$$\frac{(x-u)^T(y-u)}{\|x-u\|_2} = \|y-u\|_2 \cos\theta = \mathrm{dist}(y,H) \le \mathrm{dist}(y,C)$$

as desired

Using the claim, we have for any y

$$\begin{aligned} \mathrm{dist}(y,C) &\ge \frac{(x-u)^T(y-x+x-u)}{\|x-u\|_2} \\ &= \|x-u\|_2 + \left(\frac{x-u}{\|x-u\|_2}\right)^T (y-x) \end{aligned}$$

Hence $g = (x-u)/\|x-u\|_2$ is a subgradient of $\mathrm{dist}(x,C)$ at x

References and further reading

- S. Boyd, Lecture notes for EE 264B, Stanford University, Spring 2010-2011
- R. T. Rockafellar (1970), "Convex analysis", Chapters 23–25
- L. Vandenberghe, Lecture notes for EE 236C, UCLA, Spring 2011-2012

KKT conditions and Duality

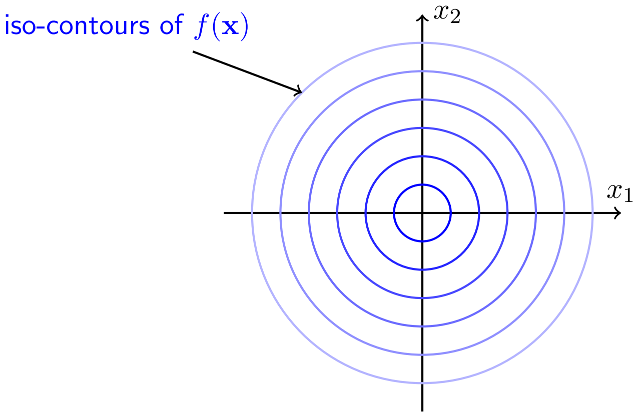
March 10, 2020

Want to solve this constrained optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^2} f(\mathbf{x}) = \min_{\mathbf{x} \in \mathbb{R}^2} .4\,(x_1^2 + x_2^2)$$

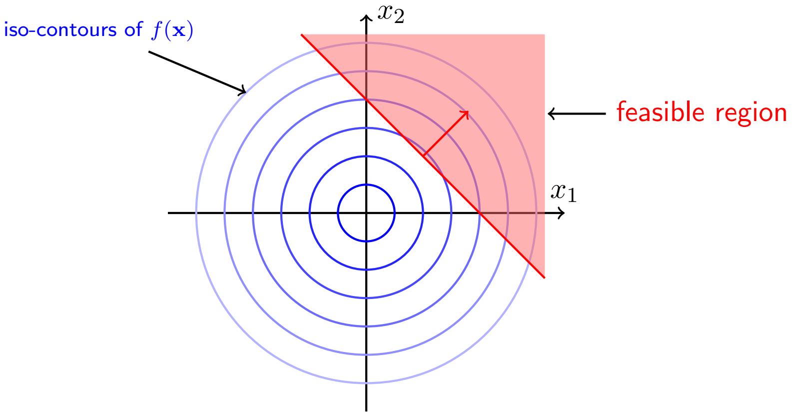
subject to

$$g(\mathbf{x}) = 2 - x_1 - x_2 \leq 0$$



$$f(\mathbf{x}) = .4\,(x_1^2 + x_2^2)$$

Tutorial example - Constraint



$$g(\mathbf{x}) = 2 - x_1 - x_2 \leq 0$$

Solve this problem with Lagrange Multipliers

Can solve this constrained optimization with Lagrange multipliers:

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda\, g(\mathbf{x})$$

Solution:

The Lagrangian is

$$\mathcal{L}(\mathbf{x}, \lambda) = .4\, x_1^2 + .4\, x_2^2 + \lambda\,(2 - x_1 - x_2)$$

The KKT conditions say that at an optimum $\lambda^* \geq 0$ and

$$\frac{\partial \mathcal{L}(\mathbf{x}^*, \lambda^*)}{\partial x_1} = .8\, x_1^* - \lambda^* = 0$$
$$\frac{\partial \mathcal{L}(\mathbf{x}^*, \lambda^*)}{\partial x_2} = .8\, x_2^* - \lambda^* = 0$$
$$\frac{\partial \mathcal{L}(\mathbf{x}^*, \lambda^*)}{\partial \lambda} = 2 - x_1^* - x_2^* = 0$$

Solve this problem with Lagrange Multipliers

Can solve this constrained optimization with Lagrange multipliers:

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda\, g(\mathbf{x})$$

Solution ctd:

Find $(x_1^*, x_2^*, \lambda^*)$ which fulfill these simultaneous equations. The first two equations imply

$$x_1^* = \frac{5}{4}\lambda^*, \qquad x_2 = \frac{5}{4}\lambda^*$$

Substituting these into the last equation we get

$$8 - 5\lambda^* - 5\lambda^* = 0 \quad \Longrightarrow \lambda^* = \frac{4}{5} \leftarrow \text{greater than 0}$$

and in turn this means

$$x_1^* = \frac{5}{4}\lambda^* = 1, \qquad x_2^* = \frac{5}{4}\lambda^* = 1$$

Solve this particular problem in another way

Alternate solution:

Construct the *Lagrangian dual function*

$$q(\lambda) = \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda) = \min_{\mathbf{x}} \; (f(\mathbf{x}) + \lambda g(\mathbf{x}))$$

Find optimal value of $\mathbf{x}$ wrt $\mathcal{L}(\mathbf{x}, \lambda)$ in terms of the Lagrange multiplier:

$$x_1^* = \frac{5}{4}\lambda, \qquad\qquad x_2^* = \frac{5}{4}\lambda$$

Substitute back into the expression of $\mathcal{L}(\mathbf{x}, \lambda)$ to get

$$q(\lambda) = \frac{5}{4}\lambda^2 + \lambda\,(2 - \frac{5}{4}\lambda - \frac{5}{4}\lambda)$$

Find $\lambda \geq 0$ which maximizes $q(\lambda)$. Luckily in this case the global optimum of $q(\lambda)$ corresponds to the constrained optimum

$$\frac{\partial q(\lambda)}{\partial \lambda} = -\frac{5}{2}\lambda + 2 = 0 \quad \Longrightarrow \quad \lambda^* = \frac{4}{5} \quad \Longrightarrow \quad x_1^* = x_2^* = 1$$

Solve the same problem in another way

The Primal Problem

$$\min_{\mathbf{x}\in\mathbb{R}^2} f(\mathbf{x}) \quad \text{subject to} \quad g(\mathbf{x}) \leq 0$$

The **Lagrangian Dual Problem**

$$\max_{\lambda\in\mathbb{R}} q(\lambda) \quad \text{subject to} \quad \lambda \geq 0$$

where

$$q(\lambda) = \min_{\mathbf{x}\in\mathbb{R}^2} \left(f(\mathbf{x}) + \lambda\, g(\mathbf{x})\right)$$

is referred to as the *Lagrangian dual function*.

In general we will have multiple inequality and equality constraints. The statement of the **Primal Problem** is

$$\min_{\mathbf{x} \in X} f(\mathbf{x})$$

subject to

$$\mathbf{g}(\mathbf{x}) \leq \mathbf{0} \quad \text{and} \quad \mathbf{h}(\mathbf{x}) = \mathbf{0}$$

Lagrangian Dual Problem

$$\max_{\boldsymbol{\lambda}, \boldsymbol{\mu}} q(\boldsymbol{\lambda}, \boldsymbol{\mu}) \text{ subject to } \boldsymbol{\lambda} \geq \mathbf{0}$$

where

$$q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\mathbf{x}} \left[f(\mathbf{x}) + \boldsymbol{\lambda}^t \, \mathbf{g}(\mathbf{x}) + \boldsymbol{\mu}^t \, \mathbf{h}(\mathbf{x}) \right]$$

is the *Lagrangian dual function*.

Why ??

This dual approach is **not guaranteed to succeed**. However,

- It does for a certain class of functions
- In these cases it often leads to a simpler optimization problem.
- Particularly in the case when the dimension of $\mathbf{x}$ is much larger than the number of constraints.
- The expression of $\mathbf{x}^*$ in terms of the Lagrange multipliers may give some insight into the optimal solution i.e. the optimal separating hyper-plane found by the SVM.

Why ??

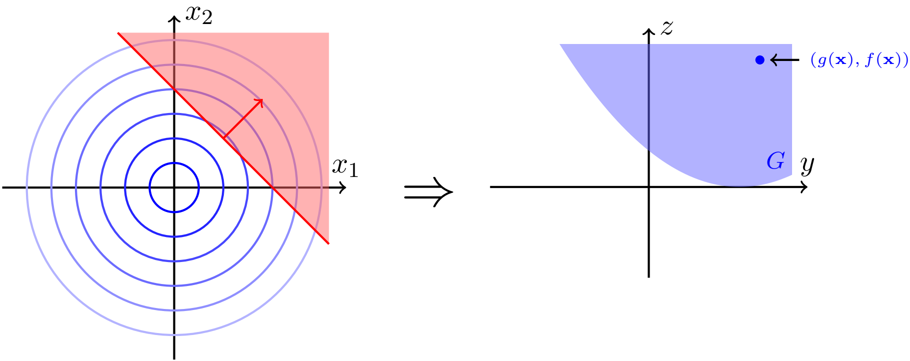
This dual approach is **not guaranteed to succeed**. However,

- It does for a certain class of functions
- In these cases it often leads to a simpler optimization problem.
- Particularly in the case when the dimension of $\mathbf{x}$ is much larger than the number of constraints.
- The expression of $\mathbf{x}^*$ in terms of the Lagrange multipliers may give some insight into the optimal solution i.e. the optimal separating hyper-plane found by the SVM.

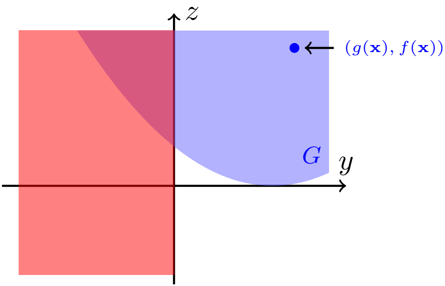
We will now focus on the geometry of the dual solution...

Geometry of the Dual Problem

Map the original problem



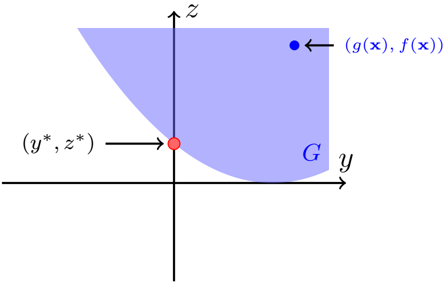
- Map each point $\mathbf{x} \in \mathbb{R}^2$ to $(g(\mathbf{x}), f(\mathbf{x})) \in \mathbb{R}^2$.
- This map defines the set
$$G = \{(y, z) \,|\, y = g(\mathbf{x}),\ z = f(\mathbf{x}) \text{ for some } \mathbf{x} \in \mathbb{R}^2\}.$$
- **Note:** $\mathcal{L}(\mathbf{x}, \lambda) = z + \lambda\, y$ for some z and y.



Define $G \subset \mathbb{R}^2$ as the image of $\mathbb{R}^2$ under the (g, f) map

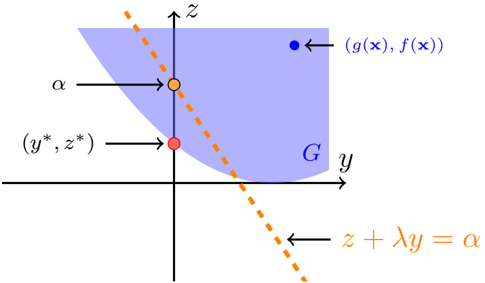
$$G = \{(y, z) \mid y = g(\mathbf{x}),\ z = f(\mathbf{x}) \text{ for some } \mathbf{x} \in \mathbb{R}^2\}$$

In this space only points with $y \leq 0$ correspond to feasible points.



- The primal problem consists in finding a point in G with $y \leq 0$ that has minimum ordinate z.
- Obviously this optimal point is (y^*, z^*).

Visualization of the Lagrangian



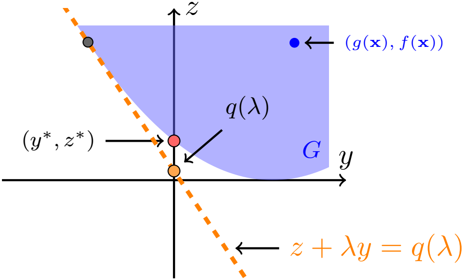
- Given a $\lambda \geq 0$, the *Lagrangian* is given by

$$\mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g(\mathbf{x}) = z + \lambda\, y$$

  with $(y, z) \in G$.

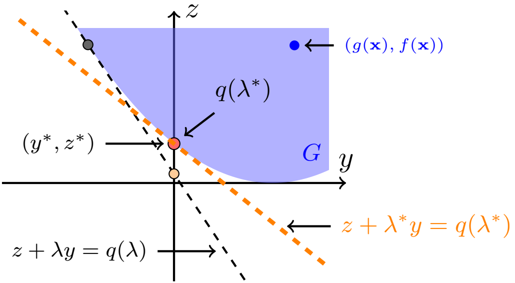
- Note $z + \lambda y = \alpha$ is the eqn of a straight line with slope $-\lambda$ that intercepts the z-axis at α.

Visualization of the Lagrangian Dual function



For a given $\lambda \geq 0$ *Lagrangian dual sub-problem* is find: $\min_{(y,z) \in G} (z + \lambda\, y)$

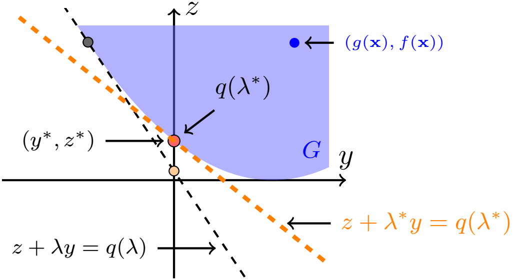
- Move the line $z + \lambda y$ in the direction $(-\lambda, -1)$ while remaining in contact with G.
- The last intercept on the z-axis obtained this way is the value of $q(\lambda)$ corresponding to the given $\lambda \geq 0$.



Finally want to find the dual optimum: $\max_{\lambda} q(\lambda)$

- the line with slope $-\lambda$ with maximal intercept, $q(\lambda)$, on the z-axis.
- This line has slope λ^* and dual optimal solution $q(\lambda^*)$.

Solving the Dual Problem



- For this problem the optimal dual objective z^* equals the optimal primal objective z^*.
- In such cases, there is **no duality gap (strong duality)**.

Properties of the Lagrangian Dual Function

Theorem

Let $D_q = \{\boldsymbol{\lambda} \,|\, q(\boldsymbol{\lambda}) > -\infty\}$ *then* $q(\boldsymbol{\lambda})$ *is concave function on* D_q.

Proof.

For any $\mathbf{x} \in X$ and $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in D_q$ and $\alpha \in (0,1)$

$$\begin{aligned}\mathcal{L}(\mathbf{x}, \alpha\boldsymbol{\lambda}_1 + (1-\alpha)\boldsymbol{\lambda}_2) &= f(\mathbf{x}) + (\alpha\boldsymbol{\lambda}_1 + (1-\alpha)\boldsymbol{\lambda}_2)^t g(\mathbf{x}) \\ &= \alpha(f(\mathbf{x}) + \boldsymbol{\lambda}_1^t g(\mathbf{x})) + (1-\alpha)(f(\mathbf{x}) + \boldsymbol{\lambda}_2^t g(\mathbf{x})) \\ &= \alpha\,\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_1) + (1-\alpha)\,\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_2).\end{aligned}$$

Take the $\min$ on both sides

$$\begin{aligned}\min_{\mathbf{x}\in X}\{\mathcal{L}(\mathbf{x}, \alpha\boldsymbol{\lambda}_1 + (1-\alpha)\boldsymbol{\lambda}_2)\} &= \min_{\mathbf{x}\in X}\{\alpha\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_1) + (1-\alpha)\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_2)\} \\ &\geq \alpha \min_{\mathbf{x}\in X}\{\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_1)\} + (1-\alpha)\min_{\mathbf{x}\in X}\{\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_2)\}\end{aligned}$$

Therefore

$$q(\alpha\boldsymbol{\lambda}_1 + (1-\alpha)\boldsymbol{\lambda}_2) \geq \alpha\, q(\boldsymbol{\lambda}_1) + (1-\alpha)\, q(\boldsymbol{\lambda}_2)$$

This implies that q is concave over D_q. □

The set of Lagrange Multipliers is convex

Theorem

Let $D_q = \{\boldsymbol{\lambda} \,|\, q(\boldsymbol{\lambda}) > -\infty\}$*. This constraint ensures valid Lagrange Multipliers exist. Then* D_q *is a convex set.*

Proof.

Let $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in D_q$. Therefore $q(\boldsymbol{\lambda}_1) > -\infty$ and $q(\boldsymbol{\lambda}_2) > -\infty$. Let $\alpha \in (0,1)$, then as q is concave

$$q(\alpha\,\boldsymbol{\lambda}_1 + (1-\alpha)\,\boldsymbol{\lambda}_2) \geq \alpha\, q(\boldsymbol{\lambda}_1) + (1-\alpha)\, q(\boldsymbol{\lambda}_2) > -\infty$$

and this implies

$$\alpha\,\boldsymbol{\lambda}_1 + (1-\alpha)\,\boldsymbol{\lambda}_2 \in D_q$$

Hence D_q is a convex set. □

Significance of these results

- The dual is always concave, irrespective of the primal problem.
- Therefore finding the **optimum of the dual function** is a **convex optimization problem**.

Weak Duality

Theorem (Weak Duality)

Let $\mathbf{x}$ *be a feasible solution,* $\mathbf{x} \in \mathcal{X}, g(\mathbf{x}) \leq 0$ *and* $h(\mathbf{x}) = 0$*, to the primal problem* P*. Let* $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ *be a feasible solution,* $\boldsymbol{\lambda} \geq 0$*, to the dual problem* D*. Then*

$$f(\mathbf{x}) \geq q(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

Proof of the Weak Duality Theorem.

Remember

$$q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf\{f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{l} \mu_i h_i(\mathbf{x}) \;:\; \mathbf{x} \in X_F\}$$

Then we have

$$\begin{aligned} q(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \inf\{f(\tilde{\mathbf{x}}) + \boldsymbol{\lambda}^t g(\tilde{\mathbf{x}}) + \boldsymbol{\mu}^t h(\tilde{\mathbf{x}}) \;:\; \tilde{\mathbf{x}} \in X_F\} \\ &\leq f(\mathbf{x}) + \boldsymbol{\lambda}^t g(\mathbf{x}) + \boldsymbol{\mu}^t h(\mathbf{x}) \\ &\leq f(\mathbf{x}) \end{aligned}$$

and the result follows. □

Corollary

Let

$$f^* = \inf\{f(\mathbf{x}) \,:\, \mathbf{x} \in X, g(\mathbf{x}) \geq 0, h(\mathbf{x}) = 0\}$$
$$q^* = \sup\{q(\boldsymbol{\lambda}, \boldsymbol{\mu}) \,:\, \boldsymbol{\lambda} \geq 0\}$$

then

$$q^* \leq f^*$$

- Thus the

  optimal value of the primal problem $\geq$ optimal value of the dual problem.
- If optimal value of the primal problem $>$ optimal value of the dual problem, then there exists a **duality gap**.

Corollary

Let

$$f^* = \inf\{f(\mathbf{x}) \,:\, \mathbf{x} \in X, g(\mathbf{x}) \geq 0, h(\mathbf{x}) = 0\}$$
$$q^* = \sup\{q(\boldsymbol{\lambda}, \boldsymbol{\mu}) \,:\, \boldsymbol{\lambda} \geq 0\}$$

then

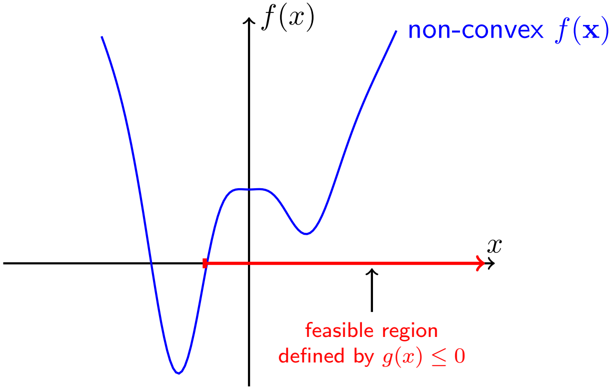
$$q^* \leq f^*$$

- Thus the

  optimal value of the primal problem $\geq$ optimal value of the dual problem.
- If optimal value of the primal problem $>$ optimal value of the dual problem, then there exists a **duality gap**.

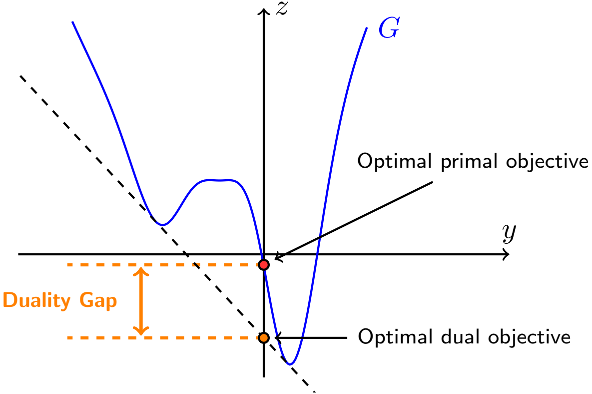
Example with a Duality Gap

Example with a non-convex objective function



- Consider the constrained optimization of this 1D non-convex objective function.
- Let's visualize $G = \{(y, z) \,|\, \exists x \in \mathbb{R} \text{ s.t. } y = g(x),\, z = f(x))\}$ and its dual solution...

Dual Solution $\leq$ Primal Solution: Have a Duality Gap



- Above is the geometric interpretation of the primal and dual problems.
- Note there exists a **duality gap** due to the nonconvexity of the set G.

Strong Duality

When does Dual Solution = Primal Solution?

The **Strong Duality Theorem** states, that if some suitable convexity conditions are satisfied, then there is no duality gap between the primal and dual optimisation problems.

Theorem (Strong Duality)

Let

- X *be a non-empty convex set in* $\mathbb{R}^n$
- $f : X \to \mathbb{R}$ *and each* $g_i : \mathbb{R}^n \to \mathbb{R}$ *(*$i = 1, \ldots, m$*) be convex,*
- *each* $h_i : \mathbb{R}^n \to \mathbb{R}$ *(*$i = 1, \ldots, l$*) be affine.*

If

- *there exists* $\hat{\mathbf{x}} \in X$ *such that* $g(\hat{\mathbf{x}}) < 0$ *and*
- $\mathbf{0} \in \operatorname{int}(\mathbf{h}(X))$ *where* $\mathbf{h}(X) = \{\mathbf{h}(\mathbf{x}) \,:\, \mathbf{x} \in X\}$.

then

$$\inf\{f(\mathbf{x}) \,:\, \mathbf{x} \in X, g(\mathbf{x}) \leq 0, h(\mathbf{x}) = 0\} = \sup\{q(\boldsymbol{\lambda}, \boldsymbol{\mu}) \,:\, \boldsymbol{\lambda} \geq \mathbf{0}\}$$

where $q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf\{f(\mathbf{x}) + \boldsymbol{\lambda}^t \mathbf{g}(\mathbf{x}) + \boldsymbol{\mu}^t \mathbf{h}(\mathbf{x}) \,:\, \mathbf{x} \in X\}$.

Theorem (Strong Duality ctd)

Furthermore, if

$$\inf\{f(\mathbf{x}) : \mathbf{x} \in X, g(\mathbf{x}) \leq 0, h(\mathbf{x}) = 0\} > -\infty$$

then the

$$\sup\{q(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\lambda} \geq 0\}$$

is achieved at $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ *with* $\boldsymbol{\lambda}^* \geq 0$*. If the* $\inf$ *is achieved at* $\mathbf{x}^*$ *then*

$$(\boldsymbol{\lambda}^*)^t \mathbf{g}(\mathbf{x}^*) = 0$$

Exercise 2.18 a. In the above proof the following property is used: if $S \subset \mathbb{R}^n$ is compact, then its convex hull co S is compact. Prove this, using the following result of Carathéodory: in $\mathbb{R}^n$ every convex combination x of $p \geq n+1$ points $x_1, \ldots, x_p$ (i.e., $x = \sum_1^p \alpha_i x_i$ for $\alpha_i \geq 0$ and $\sum_1^p \alpha_i = 1$) can also be written as a convex combination of at most $n+1$ points $x_{i_1}, \ldots, x_{i_{n+1}} \subset \{x_1, \ldots, x_p\}$.
b. Give an example of a closed set $S \subset \mathbb{R}^n$ for which co S is *not* closed (conclusion: in the above proof it is essential to work with compactness).

Exercise 2.19 Let $f(x) := |x|$ on $S := \mathbb{R}$. Then $\partial f(0) = [-1, 1]$ (by Exercise 2.16(b) for $n = 1$). Demonstrate how this result can also be derived from Theorem 2.17.

Exercise 2.20 Show by means of an example that in Theorem 2.17 it is essential to have $x_0 \in \cap_i \text{int dom } f_i$.

3 The Kuhn-Tucker theorem for convex programming

We use the results of the previous section to derive the celebrated Kuhn-Tucker theorem for convex programming. Unlike its counterparts in section 4 of [1], this theorem gives necessary *and* sufficient conditions for optimality for the standard convex programming problem. First we discuss the situation with inequality constraints only.

Theorem 3.1 (Kuhn-Tucker – no equality constraints) *Let $f, g_1, \cdots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ be convex functions and let $S \subset \mathbb{R}^n$ be a convex set. Consider the* convex programming problem

$$(P) \quad \inf_{x \in S} \{f(x) : g_1(x) \leq 0, \cdots, g_m(x) \leq 0\}.$$

Let $\bar{x}$ be a feasible point of (P); denote by $I(\bar{x})$ the set of all $i \in \{1, \cdots, m\}$ for which $g_i(\bar{x}) = 0$.

(i) $\bar{x}$ is an optimal solution of (P) if there exist vectors of multipliers $\bar{u} := (\bar{u}_1, \cdots, \bar{u}_m) \in \mathbb{R}^m_+$ *and $\bar{\eta} \in \mathbb{R}^n$ such that the following three relationships hold:*

$$\bar{u}_i g_i(\bar{x}) = 0 \text{ for } i = 1, \cdots, m \quad \textbf{(complementary slackness)},$$

$$0 \in \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + \bar{\eta} \quad \textbf{(normal Lagrange inclusion)},$$

$$\bar{\eta}^t (x - \bar{x}) \leq 0 \text{ for all } x \in S \quad \textbf{(obtuse angle property)}.$$

(ii) Conversely, if $\bar{x}$ is an optimal solution of (P) and if $\bar{x} \in$ int dom $f \cap \cap_{i \in I(\bar{x})}$int dom g_i, then there exist multipliers $\bar{u}_0 \in \{0, 1\}$, $\bar{u} \in \mathbb{R}^m_+$, $(\bar{u}_0, \bar{u}) \neq (0, 0)$, and $\bar{\eta} \in \mathbb{R}^n$ such that the complementary slackness relationship and obtuse angle property of part (i) hold, as well as the following:

$$0 \in \bar{u}_0 \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + \bar{\eta} \quad \textbf{(Lagrange inclusion)}.$$

Here the normal *case is said to occur when* $\bar{u}_0 = 1$ *and the* abnormal *case when* $\bar{u}_0 = 0$.

Remark 3.2 (minimum principle) *By Theorem 2.9, the normal Lagrange inclusion in Theorem 3.1 implies*

$$-\bar{\eta} \in \partial(f + \sum_{i \in I(\bar{x})} \bar{u}_i g_i)(\bar{x}).$$

So by Theorem 2.10 and Remark 2.11 it follows that

$$\bar{x} \in \text{argmin}_{x \in S}[f(x) + \sum_{i \in I(\bar{x})} \bar{u}_i g_i(x)] \textbf{(minimum principle)}.$$

Likewise, under the additional condition $\text{dom } f \cap \cap_{i \in I(\bar{x})} \text{int dom } g_i \neq \emptyset$*, this minimum principle implies the normal Lagrange inclusion by the converse parts of Theorem 2.10/Remark 2.11 and Theorem 2.9.*

Remark 3.3 (Slater's constraint qualification) *The following* Slater constraint qualification *guarantees normality: Suppose that there exists* $\tilde{x} \in S$ *such that* $g_i(\tilde{x}) < 0$ *for* $i = 1, \cdots, m$*. Then in part (ii) of Theorem 3.1 we have the normal case* $\bar{u}_0 = 1$*.*

Indeed, suppose we had $\bar{u}_0 = 0$*. For* $\bar{u}_0 = 0$ *instead of* $\bar{u}_0 = 1$ *the proof of the minimum principle in Remark 3.2 can be mimicked and gives*

$$\sum_{i=1}^{m} \bar{u}_i g_i(\bar{x}) \leq \sum_{i=1}^{m} \bar{u}_i g_i(\tilde{x}).$$

Since $(\bar{u}_1, \cdots, \bar{u}_m) \neq (0, \cdots, 0)$*, this gives* $\sum_{i=1}^{m} \bar{u}_i g_i(\bar{x}) < 0$*, in contradiction to complementary slackness.*

Proof of Theorem 3.1. Let us write $I := I(\bar{x})$. (*i*) By Remark 3.2 the minimum principle holds, i.e., for any $x \in S$ we have

$$f(x) + \sum_{i \in I} \bar{u}_i g_i(x) \geq f(\bar{x})$$

(observe that $\sum_{i \in I} \bar{u}_i g_i(\bar{x}) = 0$ by complementary slackness). Hence, for any *feasible* $x \in S$ we have

$$f(x) \geq f(x) + \sum_{i \in I} \bar{u}_i g_i(x) \geq f(\bar{x}),$$

by nonnegativity of the multipliers. Clearly, this proves optimality of $\bar{x}$.

(*ii*) Consider the auxiliary optimization problem

$$(P') \quad \inf_{x \in S} \phi(x),$$

where $\phi(x) := \max[f(x) - f(\bar{x}), \max_{1 \leq i \leq m} g_i(x)]$. Since $\bar{x}$ is an optimal solution of (P), it is not hard to see that $\bar{x}$ is also an optimal solution of (P') (observe that $\phi(\bar{x}) = 0$

and that $x \in S$ is feasible if and only if $\max_{1 \le i \le m} g_i(x) \le 0$). By Theorem 2.10 and Remark 2.11 there exists $\bar{\eta}$ in $\mathbb{R}^n$ such that $\bar{\eta}$ has the obtuse angle property and $-\bar{\eta} \in \partial\phi(\bar{x})$. By Theorem 2.17 this gives

$$-\bar{\eta} \in \partial\phi(\bar{x}) = \mathrm{co}(\partial f(\bar{x}) \cup \cup_{i \in I} \partial g_i(\bar{x})).$$

Since subdifferentials are convex, we get the existence of $(u_0, \xi_0) \in \mathbb{R}_+ \times \partial f(\bar{x})$ and $(u_i, \xi_i) \in \mathbb{R}_+ \times \partial g_i(\bar{x})$, $i \in I$, such that $\sum_{i \in \{0\} \cup I} u_i = 1$ and

$$-\bar{\eta} = \sum_{i \in \{0\} \cup I} u_i \xi_i.$$

In case $u_0 = 0$, we are done by setting $\bar{u}_i := u_i$ for $i \in \{0\} \cup I$ and $\bar{u}_i := 0$ otherwise. Observe that in this case $(\bar{u}_1, \cdots, \bar{u}_m) \neq (0, \cdots, 0)$ by $\sum_{i \in I} u_i = 1$. In case $u_0 \neq 0$, we know that $u_0 > 0$, so we can set $\bar{u}_i := u_i/u_0$ for $i \in \{0\} \cup I$ and $\bar{u}_i := 0$ otherwise. QED

Example 3.4 Consider the following optimization problem:

$$(P) \text{ minimize } (x_1 - \frac{9}{4})^2 + (x_2 - 2)^2$$

over all $(x_1, x_2) \in \mathbb{R}^2_+$ such that

$$\begin{aligned} x_1^2 - x_2 &\le 0 \\ x_1 + x_2 - 6 &\le 0 \\ -x_1 + 1 &\le 0 \end{aligned}$$

Since Slater's constraint qualification clearly holds, we get that a feasible point $(\bar{x}_1, \bar{x}_2)$ is optimal if and only if there exists $(\bar{u}_1, \bar{u}_2, \bar{u}_3) \in \mathbb{R}^3_+$ such that

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2(\bar{x}_1 - \frac{9}{4}) \\ 2(\bar{x}_2 - 2) \end{pmatrix} + \bar{u}_1 \begin{pmatrix} 2\bar{x}_1 \\ -1 \end{pmatrix} + \bar{u}_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \bar{u}_3 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \begin{pmatrix} \bar{\eta}_1 \\ \bar{\eta}_2 \end{pmatrix}$$

for some $\bar{\eta} := (\bar{\eta}_1, \bar{\eta}_2)^t$ with

$$\bar{\eta}^t (x - \bar{x}) \le 0 \text{ for all } x \in \mathbb{R}^2_+$$

and such that

$$\begin{aligned} \bar{u}_1(\bar{x}_1^2 - \bar{x}_2) &= 0 \\ \bar{u}_2(\bar{x}_1 + \bar{x}_2 - 6) &= 0 \\ \bar{u}_3(-\bar{x}_1 + 1) &= 0 \end{aligned}$$

Let us first deal with $\bar{\eta}$: observe that the above obtuse angle property forces $\bar{\eta}_1$ and $\bar{\eta}_2$ to be nonpositive, and $\bar{x}_i > 0$ even implies $\bar{\eta}_i = 0$ for $i = 1, 2$ (this can be seen as a form of complementarity). Since $\bar{x}_1 \ge 1$, this means $\bar{\eta}_1 = 0$. Also, $\bar{x}_2 = 0$ stands no chance, because it would mean $\bar{x}_1^2 \leq 0$. Hence, $\bar{\eta} = 0$. We now distinguish the following possibilities for the set $I := I(\bar{x})$:

Case 1 ($I = \emptyset$): By complementary slackness, $\bar{u}_1 = \bar{u}_2 = \bar{u}_3 = 0$, so the Lagrange inclusion gives $\bar{x}_1 = 9/4$, $\bar{x}_2 = 2$, which violates the first constraint ($(9/4)^2 \not\leq 2$).

Case 2 ($I = \{1\}$): By complementary slackness, $\bar{u}_2 = \bar{u}_3 = 0$. The Lagrange inclusion gives $\bar{x}_1 = \frac{9}{4}(1 + \bar{u}_1)^{-1}$, $\bar{x}_2 = \bar{u}_1/2 + 2$, so, since $\bar{x}_1^2 = \bar{x}_2$, by definition of I, we obtain the equation $\bar{u}_1^3 + 6\bar{u}_1^2 + 9\bar{u}_1 = 49/8$, which has $\bar{u}_1 = 1/2$ as its only solution. It follows then that $\bar{x} = (3/2, 9/4)^t$.

At this stage we can already stop: Theorem 3.1(i) guarantees that, in fact, $\bar{x} = (3/2, 9/4)^t$ is an optimal solution of (P). Moreover, since the objective function $(x_1, x_2) \mapsto (x_1 - \frac{9}{4})^2 + (x_2 - 2)^2$ is *strictly* convex, it follows that any optimal solution of (P) must be unique. So $\bar{x} = (3/2, 9/4)^t$ is the unique optimal solution of (P).

Exercise 3.1 Consider the optimization problem

$$(P) \quad \sup_{(\xi_1,\xi_2) \in \mathbb{R}^2_+} \{\xi_1\xi_2 : 2\xi_1 + 3\xi_2 \leq 5\}.$$

Solve this problem using Theorem 3.1. *Hint:* The set of optimal solutions does not change if we apply a monotone transformation to the objective function. So one can use $f(\xi_1, \xi_2) := \sqrt{\xi_1\xi_2}$ to ensure convexity (see Exercise 2.11).

Exercise 3.2 Let $a_i > 0$, $i = 1, \ldots, n$ and let $p \geq 1$. Consider the optimization problem

$$(P) \quad \text{maximize} \sum_{i=1}^{n} a_i\xi_i \text{ over } (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n$$

subject to $g(\xi) := \sum_{i=1}^n |\xi_i|^p = 1$.

a. Show that if the constraint $\sum_{i=1}^n |\xi_i|^p = 1$ is replaced by $\sum_{i=1}^n |\xi_i|^p \leq 1$, then this results in exactly the same optimal solutions.

b. Prove that $g : \mathbb{R}^n \to \mathbb{R}$, as defined above, is convex. Prove also that g is in fact strictly convex if $p > 1$.

c. Apply Theorem 3.1 to determine the optimal solutions of (P). *Hint:* Treat the cases $p = 1$ and $p > 1$ separately.

d. Derive from the result obtained in part (c) for $p > 1$ the following famous *Hölder inequality*, which is an extension of the Cauchy-Schwarz inequality: $|\sum_i a_i\xi_i| \leq (\sum_i a_i^q)^{1/q}(\sum_i |\xi_i|^p)^{1/p}$ for all $(\xi_1, \ldots, \xi_n) \in \mathbb{R}^n$. Here q is defined by $q := p/(p-1)$.

Corollary 3.5 (Kuhn-Tucker – general case) *Let $f, g_1, \cdots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ be convex functions, let $S \subset \mathbb{R}^n$ be a convex set. Also, let A be a $p \times n$-matrix and let $b \in \mathbb{R}^p$. Define $L := \{x : Ax = b\}$. Consider the* convex programming problem

$$(P) \quad \inf_{x \in S} \{f(x) : g_1(x) \leq 0, \cdots, g_m(x) \leq 0, Ax - b = 0\}.$$

Let $\bar{x}$ be a feasible point of (P); denote by $I(\bar{x})$ the set of all $i \in \{1, \cdots, m\}$ for which $g_i(\bar{x}) = 0$.

(i) $\bar{x}$ *is an optimal solution of* (P) *if there exist vectors of multipliers* $\bar{u} \in \mathbb{R}^m_+$, $\bar{v} \in \mathbb{R}^p$ *and* $\bar{\eta} \in \mathbb{R}^n$ *such that the complementary slackness relationship and the obtuse angle property hold just as in Theorem 3.1(i), as well as the following version of the normal Lagrange inclusion:*

$$0 \in \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + A^t \bar{v} + \bar{\eta}.$$

(ii) Conversely, if $\bar{x}$ *is an optimal solution of* (P) *and if both* $\bar{x} \in \text{int dom } f \cap \cap_{i \in I(\bar{x})} \text{int dom } g_i$ *and* $\text{int } S \cap L \neq \emptyset$*, then there exist multipliers* $\bar{u}_0 \in \{0, 1\}$, $\bar{u} \in \mathbb{R}^m_+$, $(\bar{u}_0, \bar{u}) \neq (0, 0)$*, and* $\bar{v} \in \mathbb{R}^p$, $\bar{\eta} \in \mathbb{R}^n$ *such that the complementary slackness relationship and obtuse angle property of part (i) hold, as well as the following Lagrange inclusion:*

$$0 \in \bar{u}_0 \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + A^t \bar{v} + \bar{\eta}.$$

PROOF. Observe that $\partial \chi_L(\bar{x}) = \text{im } A^t$. Indeed, $\eta \in \partial \chi_L(\bar{x})$ is equivalent to $\eta^t(x - \bar{x}) \leq 0$ for all $x \in L$, i.e., to $\eta^t(x - \bar{x}) = 0$ for all $x \in \mathbb{R}^n$ with $A(x - \bar{x}) = 0$. But the latter states that η belongs to the bi-orthoplement of the linear subspace im A^t, so it belongs to im A^t itself. This proves the observation. Let us note that the above problem (P) is precisely the same problem as the one of Theorem 3.1, but with S replaced by $S' := S \cap L$. Thus, parts (i) and (ii) follow directly from Theorem 3.1, but now $\bar{\eta}$ as in Theorem 3.1 has to be replaced by an element (say η') in $\partial \chi_{S'}$. From Theorem 2.9 we know that

$$\partial \chi_{S'}(\bar{x}) = \partial \chi_S(\bar{x}) + \partial \chi_L(\bar{x}),$$

in view of the condition int $S \cap L \neq \emptyset$. Therefore, η' can be decomposed as $\eta' = \bar{\eta} + \eta$, with $\bar{\eta} \in \partial \chi_S(\bar{x})$ (this amounts to the obtuse angle property, of course), and with $\eta \in \partial \chi_L(\bar{x})$. By the above there exists $\bar{v} \in \mathbb{R}^m$ with $\eta = A^t \bar{v}$ and this finishes the proof. QED

Example 3.6 Let $c_1, \cdots, c_n$, $a_1, \cdots, a_n$ and b be positive real numbers. Consider the following optimization problem:

$$(P) \text{ minimize } \sum_{i=1}^n \frac{c_i}{x_i}$$

over all $x = (x_1, \cdots, x_n)^t \in \mathbb{R}^n_{++}$ (the strictly positive orthant) such that

$$\sum_{i=1}^n a_i x_i = b.$$

Let us try to meet the sufficient conditions of Corollary 3.5(i). Thus, we must find a feasible $\bar{x} \in \mathbb{R}^n$ and multipliers $\bar{v} \in \mathbb{R}$, $\bar{\eta} \in \mathbb{R}^n$ such that

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{c_1}{\bar{x}_1^2} \\ \vdots \\ -\frac{c_n}{\bar{x}_n^2} \end{pmatrix} + \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \bar{v} + \bar{\eta}.$$

and such that the obtuse angle property holds for $\bar{\eta}$. To begin with the latter, since we seek $\bar{x}$ in the open set $S := \mathbb{R}^n_{++}$, the only $\bar{\eta}$ with the obtuse angle property is $\bar{\eta} = 0$. The above Lagrange inclusion gives $\bar{x}_i = (c_i/(\bar{v}a_i))^{1/2}$ for all i. To determine $\bar{v}$, which must certainly be positive, we use the constraint: $b = \sum_i a_i \bar{x}_i = \sum_i (a_i c_i/\bar{v})^{1/2}$, which gives $\bar{v} = (\sum_i (a_i c_i)^{1/2}/b)^2$. Thus, all conditions of Corollary 3.5(i) are seen to hold: an optimal solution of (P) is $\bar{x}$, given by

$$\bar{x}_i = \sqrt{\frac{c_i}{a_i}} \frac{b}{\sum_{j=1}^n \sqrt{a_j c_j}},$$

and it is implicit in our derivation that this solution is unique (exercise).

Remark 3.7 *By using the* relative interior *(denoted as "ri") of a convex set, i.e., the interior relative to the linear variety spanned by that set, one can obtain the following improvement of the nonempty intersection condition in Theorem 2.9: it is already enough that* ri dom $f \cap$ dom g *is nonempty. Since one can also prove that* $A(\text{ri } S) = \text{ri } A(S)$ *for any convex set* $S \subset \mathbb{R}^n$ *and any linear mapping* $A : \mathbb{R}^n \to \mathbb{R}^p$ *[2, Theorem 4.9], it follows that the nonempty intersection condition in Corollary 3.5 can be improved considerably into* ri $S \cap L \neq \emptyset$ *or,* equivalently*, into* $b \in A(\text{ri } S)$.

Exercise 3.3 In the above proof of Corollary 3.5 the fact was used that for a linear subspace M of $\mathbb{R}^n$ the following holds: let

$$M^\perp := \{x \in \mathbb{R}^n : x^t \xi = 0 \text{ for all } \xi \in M\},$$

This is a linear subspace itself (prove this), so $M^{\perp\perp} := (M^\perp)^\perp$ is well-defined. Prove that $M = M^{\perp\perp}$. *Hint:* This identity can be established by proving two inclusions; one of these is elementary and the other requires the use of projections.

Exercise 3.4 What becomes of Corollary 3.5 in the situation where there are no inequality constraints (i.e., just equality constraints)? Derive this version.

Exercise 3.5 Use Corollary 3.5 to prove the following famous theorem of Farkas. Let A be a $p \times n$-matrix and let $c \in \mathbb{R}^n$. Then precisely one of the following is true:

$$(1)\ \exists_{x \in \mathbb{R}^n} Ax \leq 0 \text{ (componentwise) and } c^t x > 0, \ \ (2)\ \exists_{y \in \mathbb{R}^p_+} A^t y = c.$$

Hint: Show first, by elementary means, that validity of (2) implies that (1) cannot hold. Next, apply Corollary 3.5 to a suitably chosen optimization problem in order to prove that if (1) does not hold, then (2) must be true.

MATH4230 - Optimization Theory - 2019/20

Tieyong Zeng
zeng@math.cuhk.edu.hk
March, 2020

Plan (March 10-11, 2020)

1. Review of subgradient

2. Duality

3. Kuhn-Tucker theorem

Theorem 3.1 (Kuhn-Tucker – no equality constraints) *Let* $f, g_1, \cdots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions and let* $S \subset \mathbb{R}^n$ *be a convex set. Consider the* convex programming problem

$$(P) \quad \inf_{x \in S} \{f(x) : g_1(x) \leq 0, \cdots, g_m(x) \leq 0\}.$$

Let $\bar{x}$ *be a feasible point of* (P)*; denote by* $I(\bar{x})$ *the set of all* $i \in \{1, \cdots, m\}$ *for which* $g_i(\bar{x}) = 0$.

Theorem 3.1 (Kuhn-Tucker – no equality constraints) *Let* $f, g_1, \cdots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions and let* $S \subset \mathbb{R}^n$ *be a convex set. Consider the* convex programming problem

$$(P) \quad \inf_{x \in S}\{f(x) : g_1(x) \leq 0, \cdots, g_m(x) \leq 0\}.$$

Let $\bar{x}$ *be a feasible point of* (P)*; denote by* $I(\bar{x})$ *the set of all* $i \in \{1, \cdots, m\}$ *for which* $g_i(\bar{x}) = 0$.

(*i*) $\bar{x}$ *is an optimal solution of* (P) *if there exist vectors of* multipliers $\bar{u} := (\bar{u}_1, \cdots, \bar{u}_m) \in \mathbb{R}^m_+$ *and* $\bar{\eta} \in \mathbb{R}^n$ *such that the following three relationships hold:*

$$\bar{u}_i g_i(\bar{x}) = 0 \text{ for } i = 1, \cdots, m \quad \textbf{(complementary slackness)},$$

$$0 \in \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + \bar{\eta} \quad \textbf{(normal Lagrange inclusion)},$$

$$\bar{\eta}^t(x - \bar{x}) \leq 0 \text{ for all } x \in S \quad \textbf{(obtuse angle property)}.$$

Theorem 3.1 (Kuhn-Tucker – no equality constraints) *Let* $f, g_1, \cdots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions and let* $S \subset \mathbb{R}^n$ *be a convex set. Consider the* convex programming problem

$$(P) \quad \inf_{x \in S} \{f(x) : g_1(x) \leq 0, \cdots, g_m(x) \leq 0\}.$$

Let $\bar{x}$ *be a feasible point of* (P)*; denote by* $I(\bar{x})$ *the set of all* $i \in \{1, \cdots, m\}$ *for which* $g_i(\bar{x}) = 0$.

(ii) Conversely, if $\bar{x}$ *is an optimal solution of* (P) *and if* $\bar{x} \in \text{int dom } f \cap \cap_{i \in I(\bar{x})} \text{int dom } g_i$*, then there exist multipliers* $\bar{u}_0 \in \{0, 1\}$*,* $\bar{u} \in \mathbb{R}^m_+$*,* $(\bar{u}_0, \bar{u}) \neq (0, 0)$*, and* $\bar{\eta} \in \mathbb{R}^n$ *such that the complementary slackness relationship and obtuse angle property of part (i) hold, as well as the following:*

$$0 \in \bar{u}_0 \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + \bar{\eta} \text{ (**Lagrange inclusion**)}.$$

Theorem 2.9 (Moreau-Rockafellar) *Let* $f, g : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions. Then for every* $x_0 \in \mathbb{R}^n$

$$\partial f(x_0) + \partial g(x_0) \subset \partial(f + g)(x_0).$$

Moreover, suppose that $\operatorname{int} \operatorname{dom} f \cap \operatorname{dom} g \neq \emptyset$. *Then for every* $x_0 \in \mathbb{R}^n$ *also*

$$\partial(f + g)(x_0) \subset \partial f(x_0) + \partial g(x_0).$$

Theorem 2.10 *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a convex function and let* $S \subset \mathbb{R}^n$ *be a nonempty convex set. Consider the optimization problem*

$$(P) \quad \inf_{x \in S} f(x).$$

Then $\bar{x} \in S$ *is an optimal solution of* (P) *if and only if there exists a subgradient* $\bar{\xi} \in \partial f(\bar{x})$ *such that*

$$\bar{\xi}^t (x - \bar{x}) \geq 0 \text{ for all } x \in S. \tag{1}$$

Here the normal *case is said to occur when* $\bar{u}_0 = 1$ *and the* abnormal *case when* $\bar{u}_0 = 0$.

Remark 3.2 (minimum principle) *By Theorem 2.9, the normal Lagrange inclusion in Theorem 3.1 implies*

$$-\bar{\eta} \in \partial(f + \sum_{i \in I(\bar{x})} \bar{u}_i g_i)(\bar{x}).$$

So by Theorem 2.10 and Remark 2.11 it follows that

$$\bar{x} \in \mathrm{argmin}_{x \in S}[f(x) + \sum_{i \in I(\bar{x})} \bar{u}_i g_i(x)] (\textbf{minimum principle}).$$

Likewise, under the additional condition $\mathrm{dom}\, f \cap \cap_{i \in I(\bar{x})} \mathrm{int}\, \mathrm{dom}\, g_i \neq \emptyset$, *this minimum principle implies the normal Lagrange inclusion by the converse parts of Theorem 2.10/Remark 2.11 and Theorem 2.9.*

Remark 3.3 (Slater's constraint qualification) *The following* Slater constraint qualification *guarantees normality: Suppose that there exists* $\tilde{x} \in S$ *such that* $g_i(\tilde{x}) < 0$ *for* $i = 1, \cdots, m$. *Then in part (ii) of Theorem 3.1 we have the normal case* $\bar{u}_0 = 1$.

Indeed, suppose we had $\bar{u}_0 = 0$. *For* $\bar{u}_0 = 0$ *instead of* $\bar{u}_0 = 1$ *the proof of the minimum principle in Remark 3.2 can be mimicked and gives*

$$\sum_{i=1}^{m} \bar{u}_i g_i(\bar{x}) \leq \sum_{i=1}^{m} \bar{u}_i g_i(\tilde{x}).$$

Since $(\bar{u}_1, \cdots, \bar{u}_m) \neq (0, \cdots, 0)$, *this gives* $\sum_{i=1}^{m} \bar{u}_i g_i(\bar{x}) < 0$, *in contradiction to complementary slackness.*

Theorem 3.1 (Kuhn-Tucker – no equality constraints) *Let $f, g_1, \cdots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ be convex functions and let $S \subset \mathbb{R}^n$ be a convex set. Consider the* convex programming problem

$$(P) \quad \inf_{x \in S} \{f(x) : g_1(x) \leq 0, \cdots, g_m(x) \leq 0\}.$$

Let $\bar{x}$ be a feasible point of (P); denote by $I(\bar{x})$ the set of all $i \in \{1, \cdots, m\}$ for which $g_i(\bar{x}) = 0$.

(i) $\bar{x}$ is an optimal solution of (P) if there exist vectors of multipliers *$\bar{u} := (\bar{u}_1, \cdots, \bar{u}_m) \in \mathbb{R}^m_+$ and $\bar{\eta} \in \mathbb{R}^n$ such that the following three relationships hold:*

$$\bar{u}_i g_i(\bar{x}) = 0 \text{ for } i = 1, \cdots, m \quad \textbf{(complementary slackness)},$$

$$0 \in \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + \bar{\eta} \quad \textbf{(normal Lagrange inclusion)},$$

$$\bar{\eta}^t (x - \bar{x}) \leq 0 \text{ for all } x \in S \quad \textbf{(obtuse angle property)}.$$

PROOF OF THEOREM 3.1. Let us write $I := I(\bar{x})$. (i) By Remark 3.2 the minimum principle holds, i.e., for any $x \in S$ we have

$$f(x) + \sum_{i \in I} \bar{u}_i g_i(x) \geq f(\bar{x})$$

(observe that $\sum_{i \in I} \bar{u}_i g_i(\bar{x}) = 0$ by complementary slackness). Hence, for any *feasible* $x \in S$ we have

$$f(x) \geq f(x) + \sum_{i \in I} \bar{u}_i g_i(x) \geq f(\bar{x}),$$

by nonnegativity of the multipliers. Clearly, this proves optimality of $\bar{x}$.

Theorem 2.17 (Dubovitskii-Milyutin) *Let* $f_1, \cdots, f_m : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions and let* x_0 *be a point in* $\cap_{i=1}^m \text{int dom } f_i$. *Let* $f : \mathbb{R}^n \to (-\infty, +\infty]$ *be given by*

$$f(x) := \max_{1 \le i \le m} f_i(x)$$

and let $I(x_0)$ *be the (nonempty) set of all* $i \in \{1, \cdots, m\}$ *for which* $f_i(x_0) = f(x_0)$. *Then*

$$\partial f(x_0) = \text{co } \cup_{i \in I(x_0)} \partial f_i(x_0).$$

Theorem 3.1 (Kuhn-Tucker – no equality constraints) *Let* $f, g_1, \cdots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions and let* $S \subset \mathbb{R}^n$ *be a convex set. Consider the* convex programming problem

$$(P) \quad \inf_{x \in S} \{f(x) : g_1(x) \leq 0, \cdots, g_m(x) \leq 0\}.$$

Let $\bar{x}$ *be a feasible point of* (P)*; denote by* $I(\bar{x})$ *the set of all* $i \in \{1, \cdots, m\}$ *for which* $g_i(\bar{x}) = 0$.

(ii) Conversely, if $\bar{x}$ *is an optimal solution of* (P) *and if* $\bar{x} \in \text{int dom } f \cap \cap_{i \in I(\bar{x})} \text{int dom } g_i$*, then there exist multipliers* $\bar{u}_0 \in \{0, 1\}$, $\bar{u} \in \mathbb{R}^m_+$, $(\bar{u}_0, \bar{u}) \neq (0, 0)$, *and* $\bar{\eta} \in \mathbb{R}^n$ *such that the complementary slackness relationship and obtuse angle property of part (i) hold, as well as the following:*

$$0 \in \bar{u}_0 \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + \bar{\eta} \ \textbf{(Lagrange inclusion)}.$$

(*ii*) Consider the auxiliary optimization problem

$$(P') \ \inf_{x \in S} \phi(x),$$

where $\phi(x) := \max[f(x) - f(\bar{x}), \max_{1 \le i \le m} g_i(x)]$. Since $\bar{x}$ is an optimal solution of (P), it is not hard to see that $\bar{x}$ is also an optimal solution of (P') (observe that $\phi(\bar{x}) = 0$ and that $x \in S$ is feasible if and only if $\max_{1 \le i \le m} g_i(x) \le 0$). By Theorem 2.10 and Remark 2.11 there exists $\bar{\eta}$ in $\mathbb{R}^n$ such that $\bar{\eta}$ has the obtuse angle property and $-\bar{\eta} \in \partial \phi(\bar{x})$. By Theorem 2.17 this gives

$$-\bar{\eta} \in \partial \phi(\bar{x}) = \mathrm{co}(\partial f(\bar{x}) \cup \cup_{i \in I} \partial g_i(\bar{x})).$$

$$-\bar{\eta} \in \partial\phi(\bar{x}) = \mathrm{co}(\partial f(\bar{x}) \cup \cup_{i\in I}\partial g_i(\bar{x})).$$

Since subdifferentials are convex, we get the existence of $(u_0, \xi_0) \in \mathbb{R}_+ \times \partial f(\bar{x})$ and $(u_i, \xi_i) \in \mathbb{R}_+ \times \partial g_i(\bar{x})$, $i \in I$, such that $\sum_{i\in\{0\}\cup I} u_i = 1$ and

$$-\bar{\eta} = \sum_{i\in\{0\}\cup I} u_i\xi_i.$$

In case $u_0 = 0$, we are done by setting $\bar{u}_i := u_i$ for $i \in \{0\} \cup I$ and $\bar{u}_i := 0$ otherwise. Observe that in this case $(\bar{u}_1, \cdots, \bar{u}_m) \neq (0, \cdots, 0)$ by $\sum_{i\in I} u_i = 1$. In case $u_0 \neq 0$, we know that $u_0 > 0$, so we can set $\bar{u}_i := u_i/u_0$ for $i \in \{0\} \cup I$ and $\bar{u}_i := 0$ otherwise. QED

Example 3.4 Consider the following optimization problem:

$$(P)\ \text{minimize}\ (x_1 - \frac{9}{4})^2 + (x_2 - 2)^2$$

over all $(x_1, x_2) \in \mathbb{R}^2_+$ such that

$$\begin{aligned} x_1^2 - x_2 &\leq 0 \\ x_1 + x_2 - 6 &\leq 0 \\ -x_1 + 1 &\leq 0 \end{aligned}$$

Example 3.4 Consider the following optimization problem:

$$(P) \text{ minimize } (x_1 - \frac{9}{4})^2 + (x_2 - 2)^2$$

over all $(x_1, x_2) \in \mathbb{R}^2_+$ such that

$$\begin{aligned} x_1^2 - x_2 &\leq 0 \\ x_1 + x_2 - 6 &\leq 0 \\ -x_1 + 1 &\leq 0 \end{aligned}$$

Since Slater's constraint qualification clearly holds, we get that a feasible point $(\bar{x}_1, \bar{x}_2)$ is optimal if and only if there exists $(\bar{u}_1, \bar{u}_2, \bar{u}_3) \in \mathbb{R}^3_+$ such that

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2(\bar{x}_1 - \frac{9}{4}) \\ 2(\bar{x}_2 - 2) \end{pmatrix} + \bar{u}_1 \begin{pmatrix} 2\bar{x}_1 \\ -1 \end{pmatrix} + \bar{u}_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \bar{u}_3 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \begin{pmatrix} \bar{\eta}_1 \\ \bar{\eta}_2 \end{pmatrix}$$

for some $\bar{\eta} := (\bar{\eta}_1, \bar{\eta}_2)^t$ with

$$\bar{\eta}^t (x - \bar{x}) \leq 0 \text{ for all } x \in \mathbb{R}^2_+$$

and such that

$$\begin{aligned} \bar{u}_1(\bar{x}_1^2 - \bar{x}_2) &= 0 \\ \bar{u}_2(\bar{x}_1 + \bar{x}_2 - 6) &= 0 \\ \bar{u}_3(-\bar{x}_1 + 1) &= 0 \end{aligned}$$

$$\bar{\eta}^t(x - \bar{x}) \leq 0 \text{ for all } x \in \mathbb{R}^2_+$$

and such that

$$\begin{aligned} x_1^2 - x_2 &\leq 0 \\ x_1 + x_2 - 6 &\leq 0 \\ -x_1 + 1 &\leq 0 \end{aligned}$$

$$\begin{aligned} \bar{u}_1(\bar{x}_1^2 - \bar{x}_2) &= 0 \\ \bar{u}_2(\bar{x}_1 + \bar{x}_2 - 6) &= 0 \\ \bar{u}_3(-\bar{x}_1 + 1) &= 0 \end{aligned}$$

Let us first deal with $\bar{\eta}$: observe that the above obtuse angle property forces $\bar{\eta}_1$ and $\bar{\eta}_2$ to be nonpositive, and $\bar{x}_i > 0$ even implies $\bar{\eta}_i = 0$ for $i = 1, 2$ (this can be seen as a form of complementarity). Since $\bar{x}_1 \geq 1$, this means $\bar{\eta}_1 = 0$. Also, $\bar{x}_2 = 0$ stands no chance, because it would mean $\bar{x}_1^2 \leq 0$. Hence, $\bar{\eta} = 0$.

Let $\bar{x}$ be a feasible point of (P); denote by $I(\bar{x})$ the set of all $i \in \{1, \cdots, m\}$ for which $g_i(\bar{x}) = 0$.

no chance, because it would mean $\bar{x}_1^2 \leq 0$. Hence, $\bar{\eta} = 0$. We now distinguish the following possibilities for the set $I := I(\bar{x})$:

Case 1 ($I = \emptyset$): By complementary slackness, $\bar{u}_1 = \bar{u}_2 = \bar{u}_3 = 0$, so the Lagrange inclusion gives $\bar{x}_1 = 9/4$, $\bar{x}_2 = 2$, which violates the first constraint $((9/4)^2 \not\leq 2)$.

$$\begin{aligned} x_1^2 - x_2 &\leq 0 \\ x_1 + x_2 - 6 &\leq 0 \\ -x_1 + 1 &\leq 0 \end{aligned}$$

$$\begin{aligned} \bar{u}_1(\bar{x}_1^2 - \bar{x}_2) &= 0 \\ \bar{u}_2(\bar{x}_1 + \bar{x}_2 - 6) &= 0 \\ \bar{u}_3(-\bar{x}_1 + 1) &= 0 \end{aligned}$$

Case 2 ($I = \{1\}$): By complementary slackness, $\bar{u}_2 = \bar{u}_3 = 0$. The Lagrange inclusion gives $\bar{x}_1 = \frac{9}{4}(1 + \bar{u}_1)^{-1}$, $\bar{x}_2 = \bar{u}_1/2 + 2$, so, since $\bar{x}_1^2 = \bar{x}_2$, by definition of I, we obtain the equation $\bar{u}_1^3 + 6\bar{u}_1^2 + 9\bar{u}_1 = 49/8$, which has $\bar{u}_1 = 1/2$ as its only solution. It follows then that $\bar{x} = (3/2, 9/4)^t$.

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2(\bar{x}_1 - \frac{9}{4}) \\ 2(\bar{x}_2 - 2) \end{pmatrix} + \bar{u}_1 \begin{pmatrix} 2\bar{x}_1 \\ -1 \end{pmatrix} + \bar{u}_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \bar{u}_3 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \begin{pmatrix} \bar{\eta}_1 \\ \bar{\eta}_2 \end{pmatrix}$$

At this stage we can already stop: Theorem 3.1(i) guarantees that, in fact, $\bar{x} = (3/2, 9/4)^t$ is an optimal solution of (P). Moreover, since the objective function $(x_1, x_2) \mapsto (x_1 - \frac{9}{4})^2 + (x_2 - 2)^2$ is *strictly* convex, it follows that any optimal solution of (P) must be unique. So $\bar{x} = (3/2, 9/4)^t$ is the unique optimal solution of (P).

Corollary 3.5 (Kuhn-Tucker – general case) *Let* $f, g_1, \cdots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions, let* $S \subset \mathbb{R}^n$ *be a convex set. Also, let* A *be a* $p \times n$*-matrix and let* $b \in \mathbb{R}^p$*. Define* $L := \{x : Ax = b\}$*. Consider the* convex programming problem

$$(P) \quad \inf_{x \in S} \{f(x) : g_1(x) \leq 0, \cdots, g_m(x) \leq 0, Ax - b = 0\}.$$

Let $\bar{x}$ *be a feasible point of* (P)*; denote by* $I(\bar{x})$ *the set of all* $i \in \{1, \cdots, m\}$ *for which* $g_i(\bar{x}) = 0$.

(i) $\bar{x}$ *is an optimal solution of* (P) *if there exist vectors of multipliers* $\bar{u} \in \mathbb{R}^m_+$, $\bar{v} \in \mathbb{R}^p$ *and* $\bar{\eta} \in \mathbb{R}^n$ *such that the complementary slackness relationship and the obtuse angle property hold just as in Theorem 3.1(i), as well as the following version of the normal Lagrange inclusion:*

$$0 \in \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + A^t \bar{v} + \bar{\eta}.$$

(ii) Conversely, if $\bar{x}$ *is an optimal solution of* (P) *and if both* $\bar{x} \in \operatorname{int} \operatorname{dom} f \cap \bigcap_{i \in I(\bar{x})} \operatorname{int} \operatorname{dom} g_i$ *and* $\operatorname{int} S \cap L \neq \emptyset$, *then there exist multipliers* $\bar{u}_0 \in \{0, 1\}$, $\bar{u} \in \mathbb{R}^m_+$, $(\bar{u}_0, \bar{u}) \neq (0, 0)$, *and* $\bar{v} \in \mathbb{R}^p$, $\bar{\eta} \in \mathbb{R}^n$ *such that the complementary slackness relationship and obtuse angle property of part (i) hold, as well as the following Lagrange inclusion:*

$$0 \in \bar{u}_0 \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + A^t \bar{v} + \bar{\eta}.$$

PROOF. Observe that $\partial\chi_L(\bar{x}) = \text{im } A^t$. Indeed, $\eta \in \partial\chi_L(\bar{x})$ is equivalent to $\eta^t(x - \bar{x}) \leq 0$ for all $x \in L$, i.e., to $\eta^t(x - \bar{x}) = 0$ for all $x \in \mathbb{R}^n$ with $A(x - \bar{x}) = 0$. But the latter states that η belongs to the bi-orthoplement of the linear subspace im A^t, so it belongs to im A^t itself. This proves the observation. Let us note that the above problem (P) is precisely the same problem as the one of Theorem 3.1, but with S replaced by $S' := S \cap L$. Thus, parts (i) and (ii) follow directly from Theorem 3.1, but now $\bar{\eta}$ as in Theorem 3.1 has to be replaced by an element (say η') in $\partial\chi_{S'}$. From Theorem 2.9 we know that

$$\partial\chi_{S'}(\bar{x}) = \partial\chi_S(\bar{x}) + \partial\chi_L(\bar{x}),$$

in view of the condition int $S \cap L \neq \emptyset$. Therefore, η' can be decomposed as $\eta' = \bar{\eta} + \eta$, with $\bar{\eta} \in \partial\chi_S(\bar{x})$ (this amounts to the obtuse angle property, of course), and with $\eta \in \partial\chi_L(\bar{x})$. By the above there exists $\bar{v} \in \mathbb{R}^m$ with $\eta = A^t\bar{v}$ and this finishes the proof. QED

https://math.stackexchange.com/questions/1205388/is-the-formula-textker-a-perp-textim-at-necessarily-true

Example 3.6 Let $c_1, \cdots, c_n$, $a_1, \cdots, a_n$ and b be positive real numbers. Consider the following optimization problem:

$$(P) \text{ minimize } \sum_{i=1}^{n} \frac{c_i}{x_i}$$

over all $x = (x_1, \cdots, x_n)^t \in \mathbb{R}^n_{++}$ (the strictly positive orthant) such that

$$\sum_{i=1}^{n} a_i x_i = b.$$

Let us try to meet the sufficient conditions of Corollary 3.5(i). Thus, we must find a feasible $\bar{x} \in \mathbb{R}^n$ and multipliers $\bar{v} \in \mathbb{R}$, $\bar{\eta} \in \mathbb{R}^n$ such that

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{c_1}{\bar{x}_1^2} \\ \vdots \\ -\frac{c_n}{\bar{x}_n^2} \end{pmatrix} + \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \bar{v} + \bar{\eta}.$$

and such that the obtuse angle property holds for $\bar{\eta}$. To begin with the latter, since we seek $\bar{x}$ in the open set $S := \mathbb{R}^n_{++}$, the only $\bar{\eta}$ with the obtuse angle property is $\bar{\eta} = 0$. The above Lagrange inclusion gives $\bar{x}_i = (c_i/(\bar{v}a_i))^{1/2}$ for all i. To determine $\bar{v}$, which must certainly be positive, we use the constraint: $b = \sum_i a_i\bar{x}_i = \sum_i (a_i c_i/\bar{v})^{1/2}$, which gives $\bar{v} = (\sum_i (a_i c_i)^{1/2}/b)^2$. Thus, all conditions of Corollary 3.5(*i*) are seen to hold: an optimal solution of (P) is $\bar{x}$, given by

$$\bar{x}_i = \sqrt{\frac{c_i}{a_i}} \frac{b}{\sum_{j=1}^n \sqrt{a_j c_j}},$$

and it is implicit in our derivation that this solution is unique (exercise).