Constrained Optimization

Constrained Optimization

A constrained optimization problem has the form

$$\min_{x \in \mathbb{R}^m} f(x) \qquad (\star)$$

$$\text{s.t.} \quad c_i(x) \le 0, \quad i = 1, \dots, n$$

where $f, c_1, \dots, c_n : \mathbb{R}^m \to \mathbb{R}$

Constrained optimization theory is concerned with necessary and sufficient conditions on a solution x^*, especially those that aid in solving the problem. The following is the most basic sufficiency condition and makes no assumptions on f, c_i.

Theorem (KKT Saddle-point condition)

Consider the Lagrangian

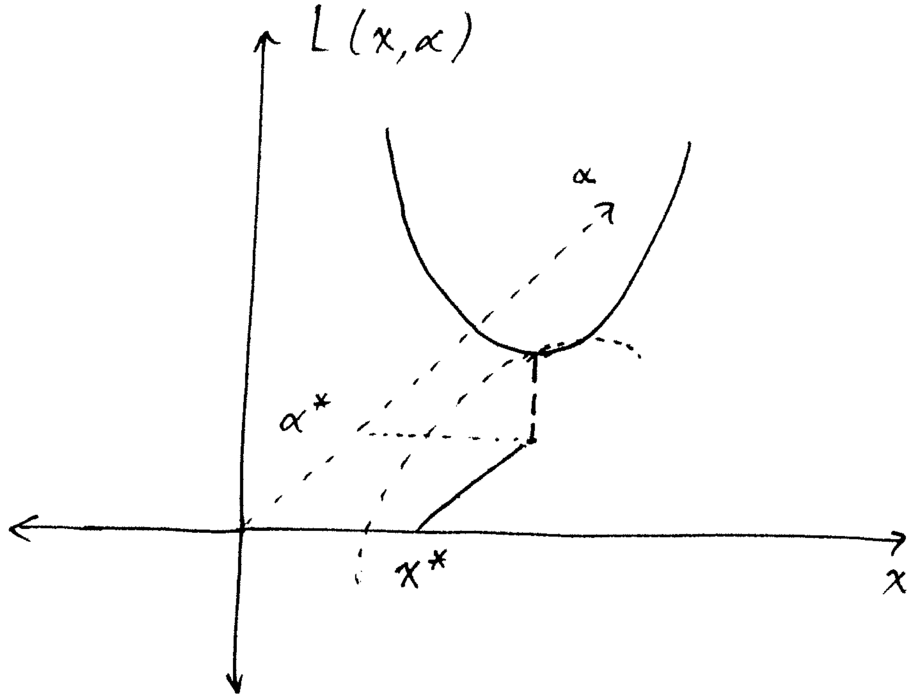
$$L(x, \alpha) := f(x) + \sum_{i=1}^{n} \alpha_i c_i(x)$$

If (x^*, α^*) is such that $\alpha_i^* \geq 0 \ \forall i$, and for any $\alpha \in [0, \infty)^n$ and $x \in \mathbb{R}^m$ satisfying $c_i(x) \le 0 \ \forall i$,

$$L(x^*, \alpha) \le L(x^*, \alpha^*) \le L(x, \alpha^*),$$

saddle point

then x^* is a solution to $(\star)$



<u>Terminology</u>: ① The parameters $\alpha_1, \dots, \alpha_n$ are called <u>Lagrange multipliers</u> ② If $c_i(x) \le 0$, $i=1,\dots,n$, we say x is <u>feasible</u>.

<u>Proof</u>: Let (x^*, α^*) be a saddle point. From $L(x^*, \alpha) \le L(x^*, \alpha^*)$ we have

$$\sum_{i=1}^{n} (\alpha_i - \alpha_i^*) c_i(x^*) \le 0.$$

This holds for any α, so take

$$\alpha = (\alpha_1^*, \dots, \alpha_{k-1}^*, \alpha_k^* + 1, \alpha_{k+1}^*, \dots, \alpha_n^*)^T$$
$$= \alpha^* + (0, \dots, 0, 1, 0, \dots, 0)^T.$$

Then

$$\sum_{i=1}^{n} (\alpha_i - \alpha_i^*) c_i(x^*) = c_k(x^*) \le 0$$

$\Longrightarrow$ x^* satisfies the constraints.

If we take

$$\alpha = (\alpha_1^*, \ldots, \alpha_{i-1}^*, 0, \alpha_{i+1}^*, \ldots, \alpha_n^*)^T \qquad \text{fix}$$

then we find

$$\alpha_i^* \cdot c_i(x^*) \geq 0, \qquad i = 1, \ldots, n.$$

Since $\alpha_i^* \geq 0$, $c_i(x^*) \leq 0$, this can only happen if

$$\alpha_i^* \cdot c_i(x^*) = 0, \qquad i = 1, \ldots, n.$$

By the second inequality we have

$$f(x^*) - f(x) = \underbrace{L(x^*, \alpha^*) - L(x, \alpha^*)}_{\leq 0} + \underbrace{\sum_{i=1}^n \alpha_i^* c_i(x)}_{\leq 0}$$

$\boxed{\text{provided } x \text{ is feasible}}$

$$\leq 0$$

QED.

When is the KKT saddle-point condition <u>necessary</u>?

An answer is that they are necessary when

- $f, c_1, \dots, c_n$ are <u>convex</u>
- one of the <u>constraint qualifications</u> in the following result is satisfied.

$\boxed{\text{conditions on } c_i}$

<u>Lemma</u>: Denote the feasible region

$$\mathcal{X} = \{x \in \mathbb{R}^m : c_i(x) \le 0, \quad i=1,\dots,n\}.$$

The following conditions are related by (i) $\Leftrightarrow$ (ii) and (iii) $\Longrightarrow$ (i), (ii):

constraint qualifications

(i) There exists $x \in \mathcal{X}$ s.t. $c_i(x) < 0$, $i=1,\dots,n$

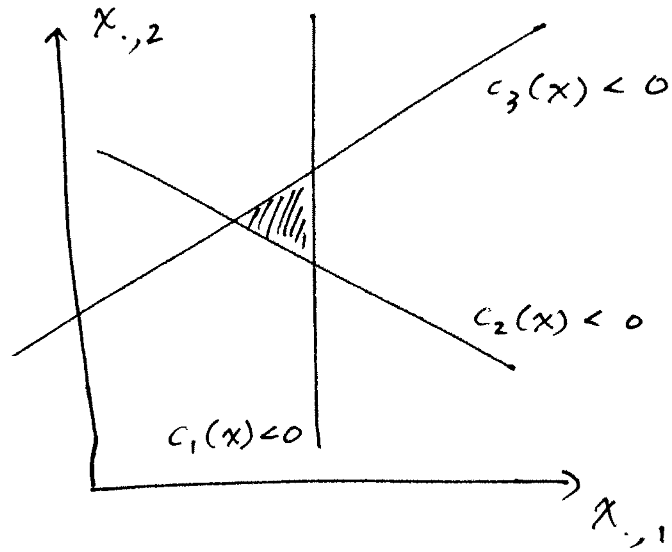
(ii) For all $\alpha \in [0,\infty)^n$, there exists $x \in \mathcal{X}$ s.t.

$$\sum_{i=1}^{n} \alpha_i c_i(x) \le 0.$$

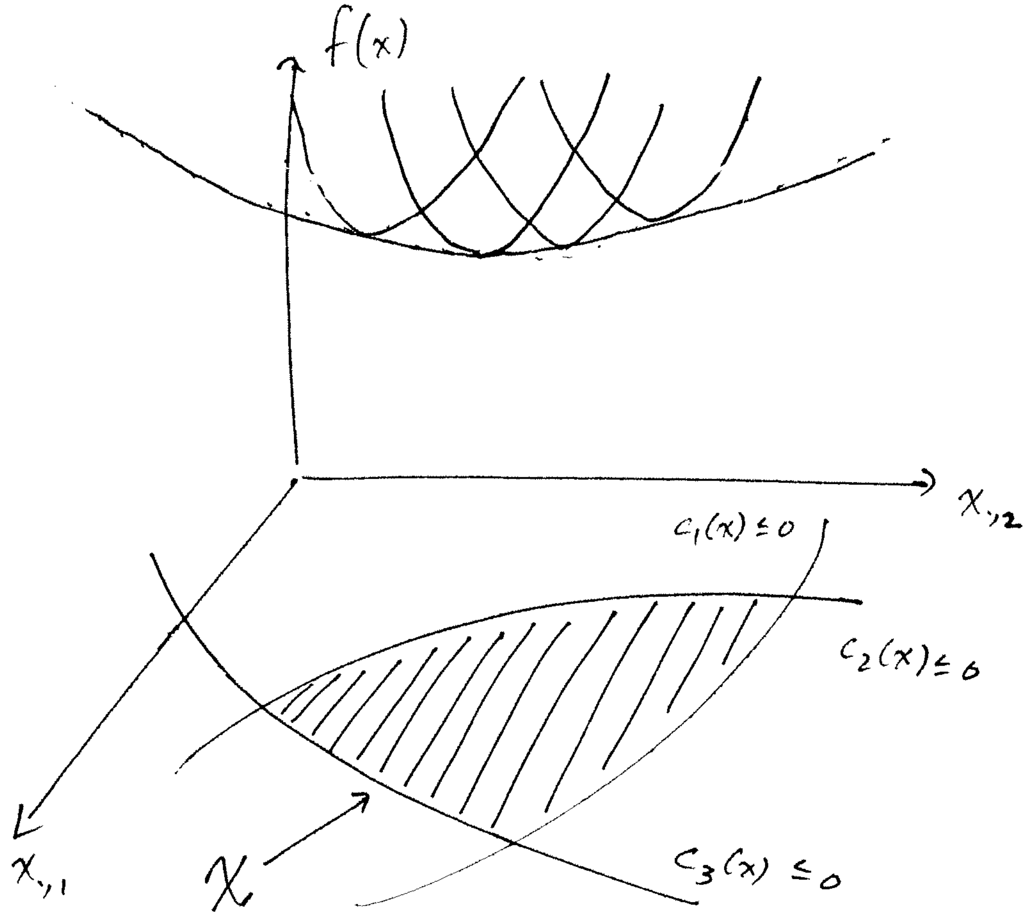
(iii) $|\mathcal{X}| \ge 2$ and there exists $x \in \mathcal{X}$ s.t. all c_i are strictly convex at x. w.r.t. $\mathcal{X}$.

(proof omitted)

Example: We will be concerned primarily with linear constraints. The first constraint qual. says that the interior of $\mathcal{X}$ is nonempty



Theorem If $f, c_1, \ldots, c_n$ are convex and one of the constraint qualifications of the preceeding Lemma holds, then the KKT saddle point condition is necessary for x^* to be optimal.



Exercise: Show that $\mathcal{X}$ is a convex <u>set</u> if $c_1, \dots, c_n$ are all convex <u>functions</u>.

Solution | Let $x_1, x_2 \in \mathcal{X}$. Must show that for any λ, $0 \le \lambda \le 1$, that $\lambda x_1 + (1-\lambda) x_2 \in \mathcal{X}$. But

$$\lambda x_1 + (1-\lambda) x_2 \in \mathcal{X} \iff c_i(\lambda x_1 + (1-\lambda) x_2) \le 0 \quad \forall i$$

Since c_i is convex, we know

$$c_i(\lambda x_1 + (1-\lambda) x_2) \le \lambda c_i(x_1) + (1-\lambda) c_i(x_2)$$
$$\le 0$$

and therefore $\lambda x_1 + (1-\lambda) x_2 \in \mathcal{X}$. ∎

Differentiable Convex Problems

When $f, c_1, \dots, c_n$ are differentiable and convex, the saddlepoint condition can be replaced by derivative-based conditions which are more useful in practice.

<u>Theorem</u> (KKT for Differentiable Convex Problems)

If $f, c_1, \ldots, c_n$ are convex and differentiable, then x^* is optimal provided $\exists\, \alpha^* \in [0, \infty)^n$ such that

- $$\frac{\partial L}{\partial x}(x^*, \alpha^*) = \frac{\partial f}{\partial x}(x^*) + \sum_{i=1}^{n} \alpha_i^* \frac{\partial c_i}{\partial x}(x^*) = 0$$

- $$c_i(x^*) \leq 0 \qquad \forall i$$

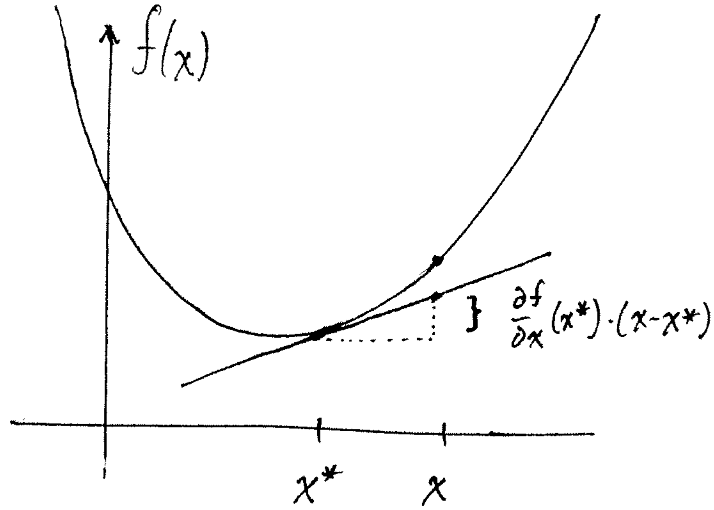
- $$\sum_{i=1}^{n} \alpha_i^* c_i(x_i^*) = 0.$$

<u>Remark</u>: The last condition implies (and is equivalent to)

$$\alpha_i^* \cdot c_i(x^*) = 0, \qquad i = 1, \ldots, n$$

Proof

Assume $x \in \mathcal{X}$ and x^*, α^* satisfies the three conditions.



Then

$$f(x) - f(x^*) \geq \frac{\partial f}{\partial x}(x^*)\cdot(x - x^*)$$

$\boxed{\text{by convexity of } f}$

$$= -\sum_{i=1}^{n} \alpha_i^* \cdot \frac{\partial c_i}{\partial x}(x^*)\cdot(x - x^*)$$

$\boxed{\text{since } \frac{\partial L}{\partial x}(x^*, \alpha^*) = 0}$

$$\geq -\sum_{i=1}^{n} \alpha_i^*\left(c_i(x) - c_i(x^*)\right)$$

$\boxed{\text{by convexity of } c_i}$

$$= -\sum_{i=1}^{n} \alpha_i^* c_i(x)$$

$\boxed{\text{by third condition}}$

$$\geq 0$$

$\boxed{\text{since } x \in \mathcal{X}}$

Thus the KKT conditions are sufficient for optimality. When are they also necessary?

<u>Theorem</u>: If $f, c_1, \dots, c_n$ are convex and differentiable, and if one of the constraint qualifications (discussed earlier) holds, then the KKT conditions of the previous result are necessary for optimality.

<u>Conclusion</u>: We can solve (A) by solving systems of equations.

Duality

Idea: Every <u>primal</u> constrained minimization problem can be converted to a <u>dual</u> constrained maximization problem (w.r.t. α) by writing down the saddle point of the Lagrangian and expressing x in terms of α.

Definition The dual of (A) is

$$\max_{(x,\alpha)} \quad L(x,\alpha) = f(x) + \sum_{i=1}^{n} \alpha_i c_i(x)$$

$$\text{s.t.} \quad \frac{\partial L}{\partial x}(x,\alpha) = 0$$

$$x \in \mathcal{X}$$

$$\alpha_i \geq 0 \quad \forall i$$

Theorem (Wolfe)

Assume

- $f, c_1, \ldots, c_n$ are convex, differentiable
- one of the constraint quals. holds

Let x^* denote a solution to the primal problem (A).

Then $\exists\, \alpha^*$ such that

- (x^*, α^*) is a solution to the dual
- $L(x^*, \alpha^*) = f(x^*)$.

Example: Linear Program

The primal linear program (LP) is

$$\min_{x} \quad c^T x$$

$$\text{s.t.} \quad Ax + d \leq 0$$

Exercise: What are the KKT sufficient conditions on (x, α) for x to be optimal?

Solution

1. $$\frac{\partial L}{\partial x}(x,\alpha) = \frac{\partial}{\partial x}\left[c^T x + \alpha^T (Ax + d)\right]$$

$$= A^T \alpha + c = 0$$

2. $Ax + d \leq 0$

3. $\alpha^T (Ax + d) = 0$

4. $\alpha \geq 0$

$\blacksquare$

The dual LP is then

$$\max_{(x,\alpha)} \quad L(x,\alpha) = c^T x + \alpha^T (Ax + d)$$

$$\text{s.t.} \quad A^T \alpha + c = 0$$

$$Ax + d \leq 0$$

$$\alpha \geq 0$$

Assume a constraint qual is satisfied. Then

$$A^T \alpha + c = 0$$

is also <u>necessary</u>. This allows us to eliminate x!

$$\Longrightarrow \quad L(x, \alpha) = c^T x + \alpha^T (Ax + d)$$

$$= -\alpha^T A x + \alpha^T A x + \alpha^T d$$

$$= d^T \alpha$$

$\Longrightarrow$ the dual LP is

$$\boxed{\begin{aligned} \max_{\alpha} \quad & d^T \alpha \\ \text{s.t.} \quad & A^T \alpha + c = 0 \\ & \alpha \geq 0. \end{aligned}}$$

<u>Trichotomy</u> For linear and convex quadratic programs exactly one of the following relates primal and dual problems:

- both feasible regions are empty
- exactly one feasible region is empty, in which case the objective function of the other is unbounded in the direction of optimization.
- both feasible regions are nonempty, in which case both problems have solutions and their extrema are equal.

<u>Exercise</u> Find the dual for the standard QP

$$\min \quad \tfrac{1}{2} x^T K x + c^T x$$

$$\text{s.t.} \quad Ax + d \le 0$$

where K is pos. def.

Solution | The saddle point condition is

$$\frac{\partial L}{\partial x}(x,\alpha) = \frac{\partial}{\partial x}\left[\frac{1}{2}x^T K x + c^T x + \alpha^T (Ax+d)\right]$$

$$= Kx + A^T\alpha + c = 0$$

The dual is

$$\max_{(x,\alpha)} \quad \frac{1}{2}x^T K x + c^T x + \alpha^T (Ax+d)$$

Assuming the feasible set has a nonempty interior (so that the constraint qual. holds), the KKT conditions are necessary. Therefore, we may plug

$$x = -K^{-1}A^T\alpha - K^{-1}c$$

into the dual.

The result is

$$L(x,\alpha) = \frac{1}{2}\left(K^{-1}A^T\alpha + K^{-1}c\right)^T K \left(K^{-1}A^T\alpha + K^{-1}c\right)$$

$$- c^T\left(K^{-1}A^T\alpha + K^{-1}c\right) - \alpha^T\left[A\cdot\left(K^{-1}A^T\alpha + K^{-1}c\right) - \right.$$

$$= -\frac{1}{2}\alpha^T\left(AK^{-1}A^T\right)\alpha + \left[d - c^TK^{-1}A^T\right]\alpha \underbrace{- \frac{1}{2}c^TK^{-1}c}_{\boxed{\text{constant term}}}$$

Ignoring the constant term the dual is

$$\max_{(x,\alpha)} \quad -\frac{1}{2}\alpha^T\left(AK^{-1}A^T\right)\alpha + \left[d - c^TK^{-1}A^T\right]\alpha$$

$$\text{s.t.} \quad Ax + d \le 0$$

$$\alpha \ge 0$$

But the first constraint can be dropped (assuming at least one feasible x exists) since x does not appear in the objective function.

We get the dual QP.

$$\boxed{\begin{array}{ll} \max\limits_{\alpha} & -\frac{1}{2}\alpha^T (AK^{-1}A^T)\alpha + [d - c^T K^{-1} A^T]\alpha \\ \\ \text{s.t.} & \alpha_i \geq 0, \quad i = 1, \ldots, n. \end{array}}$$

<u>Question</u> Why does the dual QP have fewer constraints than the dual LP?

<u>Answer</u> To simplify the dual LP we used the equation

$$A^T\alpha + c = 0$$

which <u>does not</u> involve x. To solve the dual QP we used

$$Kx + A^T\alpha + c = 0$$

which <u>does</u> involve x. In particular

$$-K(K^{-1}A^T\alpha + K^{-1}c) + A^T\alpha + c \equiv 0$$

for all α so the constraint can be eliminated.

<u>Exercise</u> Recall the QP defining the optimal margin hyperplane

$$\min_{w,b} \quad \frac{1}{2} w^T w$$

$$\text{s.t.} \quad y_i \left(w^T x_i + b \right) \geq 1 \,, \qquad i = 1, \dots, n$$

Write down the Lagrangian and KKT conditions.

Compute the dual and eliminate w, b.

<u>Hint</u>: Don't try putting the QP is standard form and applying the general result we just saw. (Why not?)

Solution | Lagrangian:

$$L(w,b,\alpha) = \frac{1}{2}w^T w + \sum_{i=1}^{n} \alpha_i \left(1 - y_i\left(w^T x_i + b\right)\right)$$

KKT conditions:

1 (a) $$\frac{\partial L}{\partial w}(w^*, b^*, \alpha^*) = w^* - \sum_{i=1}^{n} \alpha_i^* y_i x_i = 0$$

(b) $$\frac{\partial L}{\partial b}(w^*, b^*, \alpha^*) = -\sum_{i=1}^{n} \alpha_i^* y_i = 0$$

2. $$y_i\left(w^{*T} x_i + b^*\right) \geq 1 \quad \forall i$$

3. $$\alpha_i^*\left(1 - y_i\left(w^{*T} x_i + b^*\right)\right) = 0 \quad \forall i$$

4. $$\alpha_i^* \geq 0 \quad \forall i$$

Dual

$$\max_{(w,b,\alpha)} \frac{1}{2}w^T w + \sum_{i=1}^{n} \alpha_i \left(1 - y_i\left(w^T x_i + b\right)\right)$$

s.t.

$$w = \sum \alpha_i y_i x_i$$

$$0 = \sum \alpha_i y_i$$

$$y_i\left(w^T x_i + b\right) \geq 1 \qquad i = 1, \dots, n$$

$$\alpha_i \geq 0 \qquad i = 1, \dots, n$$

We may plug in $w = \sum \alpha_i y_i x_i$ and drop the constraint on w since it no longer appears in the objective. Also, b disappears because of $\sum \alpha_i y_i = 0$:

We have

$$\boxed{\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^{n} \alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^{n} \alpha_i y_i = 0 \\ & \alpha_i \geq 0, \quad i = 1, \dots, n. \end{aligned}}$$