CS 524: Introduction to Optimization
Lecture 38 : Optimality Conditions

Michael Ferris

Computer Sciences Department
University of Wisconsin-Madison

December 4, 2023

Differentiable, unconstrained optimality conditions

Theorem (FOC)

If

$$x^* \textit{ solves } \min_x f(x)$$

then

$$\nabla f(x^*) = 0$$

Conversely, if f *is convex and* x^* *satisfies* $\nabla f(x^*) = 0$ *then* x^* *solves* $\min_x f(x)$.

Lagrangians and dual problems

For an optimization problem:

$$(P) : \min f(x) \text{ s.t. } g(x) \leq 0, x \in X$$

we can define the Lagrangian function:

$$L(x, \mu) = f(x) + \mu^T g(x)$$

The primal problem (P) can be seen to be:

$$\inf_{x \in X} \sup_{\mu \geq 0} L(x, \mu)$$

and the dual functional q (always concave) is defined by:

$$q(\mu) = \inf_{x \in X} L(x, \mu)$$

so the dual problem is defined as:

$$\sup_{\mu \geq 0} q(\mu)$$

In the special case of linear programming, this leads to the dual problems outlined previously.

Example: LP

For the linear programming problem

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t. } Ax &\geq b \\ x &\geq 0 \end{aligned} \quad \equiv \quad \begin{bmatrix} A \\ I \end{bmatrix} x \geq \begin{bmatrix} b \\ 0 \end{bmatrix}$$

Lagrangian is: $L(x, p, \mu) = c^T x + p^T (b - Ax) - \mu^T x$, $X = \mathbb{R}^n$
Dual functional:

$$q(\mu) = \inf_{x \in X} L(x, \mu)$$

$$\nabla_x L(x, \mu) = c - A^T p - \mu = 0$$

Dual problem is:

$$\max_{p,\mu} p^T b \text{ s.t. } c - A^T p = \mu \geq 0, p \geq 0$$

Simplifying:

$$\max_{p} p^T b \text{ s.t. } c - A^T p \geq 0, p \geq 0$$

Constraint qualifications

In nonlinear settings, a constraint qualification is needed to ensure that a linearization of the constraint set corresponds to a "cone of feasible directions". Three popular constraint qualifications (when $X = \mathbb{R}^n$) are:

1. Polyhedral: all constraints are linear equations or inequalities
2. Slater: there exists $\hat{x}$ such that $g(\hat{x}) < 0$
3. LICQ: the gradients of the active constraints are linearly independent

KKT conditions (with differentiability)

Theorem (KKT)

Consider $f : \mathbb{R}^n \mapsto \mathbb{R}$, $g : \mathbb{R}^n \mapsto \mathbb{R}^m$:

$$(P) : \min f(x) \ s.t. \ g(x) \le 0, x \in \mathbb{R}^n$$

If x^* *solves (P) and a CQ holds, then there exists* $\mu^* \in \mathbb{R}^m$ *such that*

$$0 = \nabla_x L(x^*, \mu^*) \quad (1)$$
$$0 \le \mu^* \quad (2)$$
$$0 \le -g(x^*) \quad (3)$$
$$0 = \mu_i^* g_i(x^*), i = 1, \ldots, m \quad (4)$$

Conversely, if f *and* g *are convex and satisfy (1) - (4) then* x^* *solves (P).*

Nonnegative variables

In the special case where $g(x) = -x$ so the constraint set is just $x \geq 0$, the KKT theorem simplifies:

Theorem (KKT (nonnegative variables))

Consider $f : \mathbb{R}^n \mapsto \mathbb{R}$*:*

$$(B) : \min f(x) \ s.t. \ x \geq 0$$

If x^* *solves (B), then there exists* $\mu^* \in \mathbb{R}^m$ *such that*

$$\begin{aligned} 0 &\leq \nabla_x f(x^*), \\ 0 &\leq x^* \\ 0 &= x_i^* \nabla_x f_i(x^*), i = 1, \ldots, m \end{aligned}$$

Conversely, if f *is convex and satisfy the above then* x^* *solves (B).*

Complementarity problems

The optimality conditions of the previous slide are in the form of a complementarity problem. Such problems are defined by a function $F : \mathbb{R}^n \mapsto \mathbb{R}^n$ and seek $x^* \in \mathbb{R}^n$ satisfying:

$$\begin{aligned} 0 &\leq F(x^*), \\ 0 &\leq x^* \\ 0 &= x_i^* F_i(x^*), i = 1, \ldots, n \end{aligned}$$

Often, these conditions are written succinctly as:

$$0 \leq x^* \perp F(x^*) \geq 0$$

with the "$\perp$" being shorthand notation for the n equations setting $x_i^* F_i(x^*)$ to zero.

The optimality conditions in the nonnegative constrained optimization are then $0 \leq x^* \perp \nabla f(x^*) \geq 0$.

KKT conditions again (with differentiability)

Theorem (KKT)

Consider $f : \mathbb{R}^n \mapsto \mathbb{R}$, $g : \mathbb{R}^n \mapsto \mathbb{R}^m$:

$$(P) : \min f(x) \text{ s.t. } g(x) \leq 0, x \in \mathbb{R}^n$$

If x^* *solves (P) and a CQ holds, then there exists* $\mu^* \in \mathbb{R}^m$ *such that*

$$0 = \nabla_x L(x^*, \mu^*) \tag{5}$$

$$0 \leq \mu^* \perp -g(x^*) \geq 0 \tag{6}$$

Conversely, if f *and* g *are convex and satisfy (5) - (6) then* x^* *solves (P).*

KKT of LP: complementary slackness

For the linear programming problem

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t. } Ax &\geq b \\ x &\geq 0 \end{aligned} \quad \equiv \quad \begin{bmatrix} A \\ I \end{bmatrix} x \geq \begin{bmatrix} b \\ 0 \end{bmatrix}$$

Its KKT conditions are $(L(x, p, \mu) = c^T x + p^T(b - Ax) - \mu^T x)$

$$\begin{aligned} & 0 = c - [A^T \;\; I] \begin{bmatrix} p \\ \mu \end{bmatrix} \quad \implies \quad \mu = c - A^T p \\ & 0 \leq -b + Ax \perp p \geq 0 \\ & 0 \leq x \perp \mu \geq 0 \end{aligned}$$

or equivalently

$$\begin{aligned} 0 \leq -b + Ax \perp p \geq 0 \\ 0 \leq x \perp c - A^T p \geq 0 \end{aligned}$$

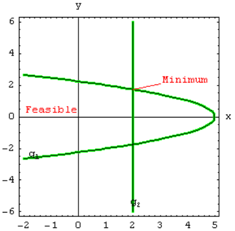
These are often termed the complementary slackness conditions of LP

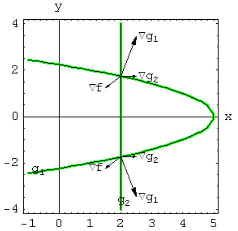
Example

For problems with two variables, it is possible to draw gradient vectors and geometrically interpret the KKT condition

$$f = -x - y \qquad \nabla f \rightarrow \begin{pmatrix} -1 \\ -1 \end{pmatrix} \qquad \nabla g_1 \rightarrow \begin{pmatrix} 1 \\ 2y \end{pmatrix} \quad \nabla g_2 \rightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$g = \{x + y^2 - 5 \le 0; x - 2 \le 0\}$$



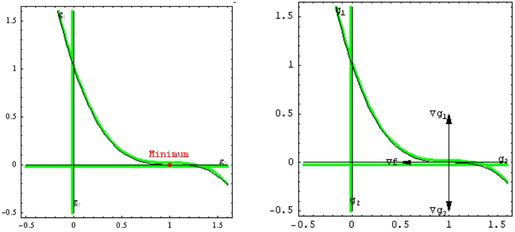


The Regularity Condition

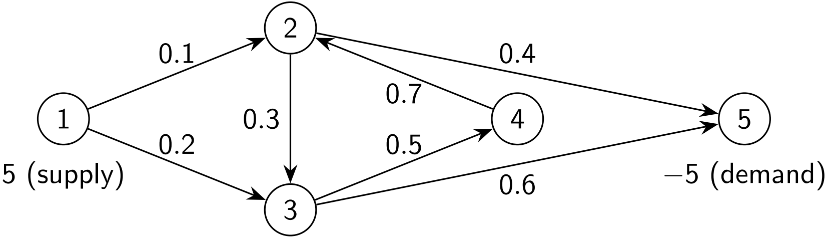
- Active inequality and all equality constraints must be linearly independent.
- The optimality conditions make sense only at points that are regular.
- The KKT condition $\nabla f(x^*) = -\sum_{i \in Active} \alpha_i \nabla g_i(x^*) + \sum_{i=1}^{p} \beta_i \nabla h_i(x^*)$ may not work for an irregular point, that can be minimum point.

$$f(x,y) = -x$$

$$g = \begin{Bmatrix} y-(1-x)^3 \le 0 \\ -x \le 0 \\ -y \le 0 \end{Bmatrix}$$



Transportation Problem (network-mcp.gms)



The Network Representation of a Transportation Problem

- $\mathcal{N}$: set of nodes in the network
- $\mathcal{A}$: set of arcs in the network
- s_i: supply of node i
- d_j: demand of node j
- x_{ij}: amount of flow through arc (i, j)
- c_{ij}: unit cost of flow associated with arc (i, j)

Constraints

For supply node i,

$$\sum_{j:(i,j)\in\mathcal{A}} x_{ij} \leq s_i$$

For demand node j,

$$\sum_{i:(i,j)\in\mathcal{A}} x_{ij} \geq d_j$$

For general node i,

$$\sum_{j:(j,i)\in\mathcal{A}} x_{ji} + s_i \geq \sum_{j:(i,j)\in\mathcal{A}} x_{ij} + d_i$$

For arc (i,j),

$$x_{ij} \geq 0$$

Linear Programming Problem

The corresponding linear programming problem is

$$\begin{aligned} \min \quad & \sum_{(i,j)\in\mathcal{A}} c_{ij}x_{ij} & & \\ \text{s.t.} \quad & \sum_{j:(j,i)\in\mathcal{A}} x_{ji} + s_i \ge \sum_{j:(i,j)\in\mathcal{A}} x_{ij} + d_i & \forall i \in \mathcal{N} \qquad & \equiv \quad Ax \ge b \\ & x_{ij} \ge 0 & \forall (i,j) \in \mathcal{A} & \end{aligned}$$

where (A is negative of node-arc incidence matrix)

$$A_{ik} = \begin{cases} -1 & \text{if } i \text{ is the tail of the } k\text{th arc} \\ +1 & \text{if } i \text{ is the head of the } k\text{th arc} \\ 0 & \text{otherwise} \end{cases}$$

$$b_i = d_i - s_i$$

Note that

$$(Ax)_i = \sum_{k\in\mathcal{A}} A_{ik}x_k = \text{inflow}(i) - \text{outflow}(i) = \sum_{j:(j,i)\in\mathcal{A}} x_{ji} - \sum_{j:(i,j)\in\mathcal{A}} x_{ij}$$

KKT Conditions

More detail: the KKT conditions are

$$0 \le Ax - b \perp p \ge 0$$
$$0 \le c - A^T p \perp x \ge 0 \quad \equiv \quad 0 \le p_i + c_{ij} - p_j \perp x_{ij} \ge 0 \quad \forall (i,j) \in \mathcal{A}$$

or

$$0 \le \begin{bmatrix} 0 & -A^T \\ A & 0 \end{bmatrix} \begin{pmatrix} x \\ p \end{pmatrix} - \begin{bmatrix} -c \\ b \end{bmatrix} \perp \begin{pmatrix} x \\ p \end{pmatrix} \ge 0$$

See `transmcp.gms` from the GAMS library to model complementarity in GAMS

Key point is in model statement / F . x / represents $F(x) \perp x$ and we can use bounds on x to generate $x \ge 0$

Shadow Price

Let p_i be the marginal retail price of goods at node i, then the optimality conditions are

$$\sum_{j:(j,i)\in\mathcal{A}} x_{ji} + s_i > \sum_{j:(i,j)\in\mathcal{A}} x_{ij} + d_i \implies p_i = 0$$

$$\sum_{j:(j,i)\in\mathcal{A}} x_{ji} + s_i = \sum_{j:(i,j)\in\mathcal{A}} x_{ij} + d_i \implies p_i \geq 0$$

$$p_i + c_{ij} \geq p_j \perp x_{ij} \geq 0 \qquad \forall (i,j) \in \mathcal{A}$$

Intuition:

- if the supply inflow$(i) + s_i$ is greater than the demand outflow$(i) + d_i$, then we are not willing to pay for even more goods
- if the price at i plus transportation cost c_{ij} exceeds the price at j, then there should be no flow through arc (i,j)

The KKT conditions are exactly these optimality conditions

Nonlinear Cost Function (`network-congestion.gms`)

To model congestion in the network, we may replace the linear cost function $c_{ij}x_{ij}$ by nonlinear cost functions like $c_{ij}x_{ij} + \beta \left(\frac{x_{ij}}{\bar{x}_{ij}} \right)^5$

The optimality conditions for nonlinear cost function $c(x)$ are the same as before except that c_{ij} is replaced by $\nabla_{x_{ij}} c(x)$:

$$
\begin{aligned}
0 \le \nabla_x c(x) - A^T p \perp x \ge 0 \\
0 \le Ax - b \perp p \ge 0
\end{aligned}
$$

or

$$
0 \le F(x,p) \perp \begin{pmatrix} x \\ p \end{pmatrix} \ge 0 \quad \text{with } F(x,p) = \begin{bmatrix} \nabla_x c(x) - A^T p \\ Ax - b \end{bmatrix}
$$