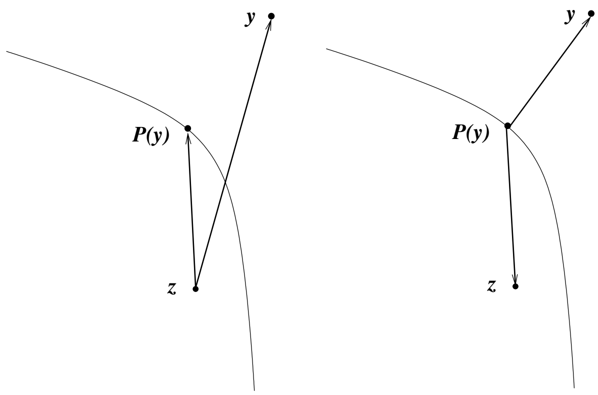

FIG. 1.1. *Lemma 1.1 illustrated: part (i) (left) and part (ii) (right).*

Lecture 33. (4/27/18; 60 min)

Optimization over Closed Convex Sets

We now discuss the more general problem of optimization over a closed convex set $\Omega \subset \mathbb{R}^n$:

$$\min f(x) \;\text{ subject to }\; x \in \Omega. \tag{1.2}$$

A point that satisfies the constraint $x \in \Omega$ is said to be *feasible*.

It is relatively easy to develop optimality conditions and the gradient projection algorithm in these general terms, then specialize it to the bound constrained case in which

$$\Omega = \{z \in \mathbb{R}^n \mid z \geq 0\}. \tag{1.3}$$

Given a closed convex set $\Omega \subset \mathbb{R}^n$, the projection operator $P : \mathbb{R}^n \to \Omega$ is defined as follows:

$$P(y) = \arg\min_{z \in \Omega} \|z - y\|_2.$$

That is, $P(y)$ is the point in Ω that is closest to y in the sense of the Euclidean norm. This operator is useful both in defining optimality conditions and in defining algorithms.

We start with a useful result about P. See Figure 1.1.

LEMMA 1.1.
(i) $(P(y) - z)^T(y - z) \geq 0$ *for all* $z \in \Omega$*, with equality if and only if* $z = P(y)$*.*
(ii) $(y - P(y))^T(z - P(y)) \leq 0$ *for all* $z \in \Omega$*.*

Proof. We prove (i) and leave (ii) as an exercise.

Let z be an arbitrary vector in Ω. We have

$$\begin{aligned} \|P(y) - y\|_2^2 &= \|P(y) - z + z - y\|_2^2 \\ &= \|P(y) - z\|_2^2 + 2(P(y) - z)^T(z - y) + \|z - y\|_2^2 \end{aligned}$$

which implies by rearrangement that

$$2(P(y) - z)^T(y - z) = \|P(y) - z\|_2^2 + \left[\|z - y\|_2^2 - \|P(y) - y\|_2^2\right]. \tag{1.4}$$

The term in brackets is nonnegative, from the definition of P. The first term on the right-hand side is trivially nonnegative, so the nonnegativity claim is proved.

If $z = P(y)$, we obviously have $(P(y) - z)^T(y - z) = 0$. If the latter condition holds, then the first term on the right-hand side of (1.4) in particular is zero, so we have $z = P(y)$. □

We now state the first-order necessary conditions for the problem (1.2).

THEOREM 1.2. *If f is continuously differentiable and x^* is a local solution of (1.2), then*

$$\nabla f(x^*)^T(x - x^*) \geq 0, \quad \textit{for all } x \in \Omega. \tag{1.5}$$

Proof. (Sketch.) Suppose that there is some x such that $\nabla f(x^*)^T(x - x^*) < 0$. Consider a step to a point $x^* + \epsilon(x - x^*)$. We have first by convexity that, since $x \in \Omega$ and $x^* \in \Omega$, that $x^* + \epsilon(x - x^*) \in \Omega$ for all $\epsilon \in [0, 1]$. In addition, we have by Taylor's theorem that

$$f(x^* + \epsilon(x - x^*)) = f(x^*) + \epsilon \nabla f(x^*)^T(x - x^*) + o(\epsilon) < f(x^*),$$

for all positive ϵ sufficiently small, since $\nabla f(x^*)^T(x - x^*)$ so the final term is dominated by the first order term for small ϵ. □

We define the normal cone as follows.

DEFINITION 1.3. *Let Ω be a closed convex set and let $x \in \Omega$. The* normal cone to Ω at x, *denoted $N_\Omega(x)$ is*

$$N_\Omega(x) := \{z \mid z^T(y - x) \leq 0 \ \textit{ for all } y \in \Omega\}.$$

Do some examples of normal cones to the closed convex sets

$$\begin{aligned}
\Omega &:= \{x \in \mathbb{R}^n \mid x \geq 0\} \quad \text{nonnegative orthant,} \\
\Omega &:= \{x \in \mathbb{R}^2 \mid \|x\|_2 \leq 1\} \quad \text{unit disk,} \\
\Omega &:= \{x \in \mathbb{R}^2 \mid \|x\|_1 \leq 1\} \quad \text{“diamond”} \\
\Omega &:= \{x \in \mathbb{R}^2 \mid \|x\|_\infty \leq 1\} \quad \text{unit box.}
\end{aligned}$$

Thus Theorem 1.2 can be stated equivalently as

$$-\nabla f(x^*) \in N_\Omega(x^*). \tag{1.6}$$

Lecture 34. (4/30/18; 60 min)

Normal Cone to the Nonnegative Orthant.. Let us verify the formula for normal cone to the nonnegative orthant $\Omega := \{x \mid x \geq 0\}$. We claim that

$$N_\Omega(x) = S := \{v \mid v_i = 0 \text{ if } x_i > 0 \text{ and } v_i \leq 0 \text{ if } x_i = 0\}. \tag{1.7}$$

We prove that $N_\Omega(x) = S$ by proving the inclusion in both directions.

$S \subset N_\Omega(x)$**.** Take any $v \in S$ and show that $v^T(z - x) \leq 0$ for all $z \geq 0$, thus verifying that $v \in N_\Omega(x)$. We have

$$v^T(z - x) = \sum_{i=1}^{n} v_i(z_i - x_i) = \sum_{i:x_i=0} v_i(z_i - x_i) = \sum_{i:x_i=0} v_i z_i \leq 0,$$

since $z_i \geq 0$, while $v_i \leq 0$ for all i with $x_i = 0$.

$N_\Omega(x) \subset S$. Here we use the fact that $v^T(z-x) \leq 0$ for all $z \geq 0$ to make careful choices of z that highlight the properties of v. First consider an index i such that $x_i = 0$. We choose z as follows:

$$z_j = x_j, \text{ for } j \neq i, \quad z_i = 1.$$

Then $z \geq 0$ and so

$$0 \geq v^T(z-x) = \sum_{j=1}^n v_j(z_j - x_j) = v_i(z_i - x_i) = v_i,$$

implying that $v_i \leq 0$ whenever $x_i = 0$. Now choose index i such that $x_i > 0$. We make two different choices of z to demonstrate that $v_i = 0$ in this case. First, set z as follows:

$$z_j = x_j, \text{ for } j \neq i, \quad z_i = 0,$$

and note that $z \geq 0$. We then have

$$0 \geq v^T(z-x) = v_i(z_i - x_i) = -v_i x_i,$$

which implies that $v_i \geq 0$, since $x_i > 0$. Now choose

$$z_j = x_j, \text{ for } j \neq i, \quad z_i = x_i + 1,$$

and note again that $z \geq 0$. We have

$$0 \geq v^T(z-x) = v_i(z_i - x_i) = v_i,$$

which implies that $v_i \leq 0$ for this i. Putting these last two facts together, we have that $x_i > 0 \Rightarrow v_i = 0$, so that $v \in N_\Omega(x) \Rightarrow v \in S$.

At this point, we have shown that $N_\Omega(x) = S$, as claimed.

From the condition $-\nabla f(x^*) \in N_\Omega(x^*)$, we have for the bound-constrained case $\Omega = \{x \,|\, x \geq 0\}$ and (1.7) that

$$x_i^* = 0 \;\;\Rightarrow\;\; [\nabla f(x^*)]_i \geq 0, \quad x_i^* > 0 \;\;\Rightarrow\;\; [\nabla f(x^*)]_i = 0.$$

Another way to state these conditions succinctly is

$$0 \leq x^* \perp \nabla f(x^*) \geq 0, \tag{1.8}$$

where $u \perp v$ indicates that $u^T v = 0$.

First-Order Conditions, Contd.. We have the following theorem concerning first-order conditions.

THEOREM 1.4. *If*

$$P(x^* - \bar{\alpha}\nabla f(x^*)) = x^*, \;\; \textit{for some } \bar{\alpha} > 0, \tag{1.9}$$

then (1.5) holds. Conversely, if (1.5) holds, then

$$P(x^* - \alpha\nabla f(x^*)) = x^*, \;\; \textit{for all } \alpha > 0.$$

Proof. Suppose first that (1.9) holds. In Lemma 1.1(ii) we set

$$y = x^* - \bar{\alpha}\nabla f(x^*), \quad P(y) = x^*,$$

and let z be any element of Ω. We then have

$$0 \geq (y - P(y))^T(z - P(y)) = (-\bar{\alpha}\nabla f(x^*))^T(z - x^*),$$

which implies that $\nabla f(x^*)^T(z - x^*) \geq 0$ for all $z \in \Omega$, proving (1.5).

Now supposed that (1.5) holds, and denote

$$x_\alpha = P(x^* - \alpha\nabla f(x^*)).$$

Setting $y = x^* - \alpha\nabla f(x^*)$, $P(y) = x_\alpha$, $z = x^*$ in Lemma 1.1, we have

$$(x^* - \alpha\nabla f(x^*) - x_\alpha)^T(x^* - x_\alpha) \leq 0,$$

which implies that

$$\|x^* - x_\alpha\|_2^2 - \alpha\nabla f(x^*)^T(x^* - x_\alpha) \leq 0. \tag{1.10}$$

By (1.5), we have that

$$-\alpha\nabla f(x^*)^T(x^* - x_\alpha) \geq 0,$$

so both terms on the left-hand side of (1.10) are nonnegative. Hence they are both zero and we have in particular that $x_\alpha = x^*$ as claimed. □

An immediate consequence of this theorem is that $P(x^* - \bar{\alpha}\nabla f(x^*)) = x^*$ for *some* $\bar{\alpha} > 0$, then $P(x^* - \alpha\nabla f(x^*)) = x^*$ for *all* $\alpha > 0$. (Why?)

Can we convert the geometric condition (1.5) into a checkable algebraic condition? Yes, if we have explicit expressions for the normal cone. The translation is easy in some cases, but less obvious in others.

Optimization with Linear Constraints.. We consider now the case in which Ω is defined by a set of linear inequalities, that is,

$$\Omega := \{x \,|\, a_i^T x \geq b_i, \quad i = 1, 2, \ldots, m\}. \tag{1.11}$$

Given a point $x^* \in \Omega$, we define the *active set* $\mathcal{A}^* \subset \{1, 2, \ldots, m\}$ to be the linear constraints that are satisfied *at equality* at the point x^*, that is,

$$\mathcal{A}^* := \{j = 1, 2, \ldots, m \,|\, a_j^T x^* = b_i\}. \tag{1.12}$$

Note that for the *inactive constraints* $i \notin \mathcal{A}^*$, we have strict inequality: $a_i^T x^* > b_i$.

The normal cone $N_\Omega(x^*)$ is the set of conic combinations of the vectors $-a_i$ whose indices are in $\mathcal{A}^*$, that is,

$$N_\Omega(x^*) = S := \{\sum_{i \in \mathcal{A}^*} \lambda_i(-a_i) \,|\, \lambda_i \geq 0, \quad i \in \mathcal{A}^*\}. \tag{1.13}$$

It is not hard to verify that $S \subset N_\Omega(x^*)$. Any point $z \in \Omega$ will have $a_i^T z \geq b_i$ and hence $a_i^T(z - x^*) \geq 0$ for all $i \in \mathcal{A}^*$, so that for $v = \sum_{i \in \mathcal{A}^*} \lambda_i(-a_i) \in \Omega$ with $\lambda_i \geq 0$ for all $i \in \mathcal{A}^*$, we have

$$v^T(z - x^*) = \sum_{i \in \mathcal{A}^*} \lambda_i(-a_i)^T(z - x^*) \leq 0,$$

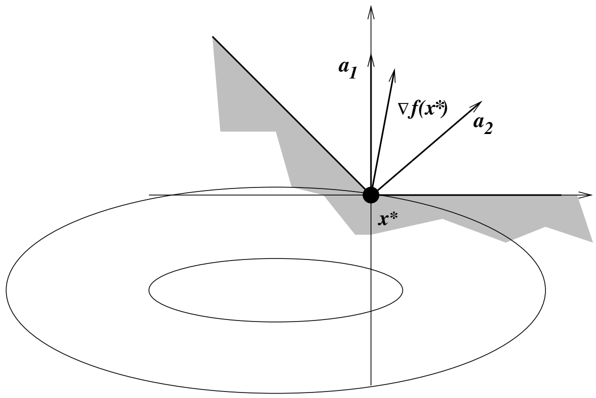

FIG. 1.2. *KKT conditions for the example, with quadratic objective centered at* $(-1,-1)^T$. *Both constraints are active at the solution.*

since all $\lambda_i \geq 0$ and $a_i^T(z - x^*) \geq 0$.

Proving that $N_\Omega(x^*) \subset S$ is more difficult.

It follows that the first-order optimality condition reduces in this case to

$$\nabla f(x^*) = \sum_{i \in \mathcal{A}^*} \lambda_i a_i, \quad \text{where } \lambda_i \geq 0 \text{ for all } i \in \mathcal{A}^*,$$

We summarize this result in a theorem.

THEOREM 1.5 (KKT conditions). *Suppose that f is continuously differentiable and x^* is a local solution of (1.2), where Ω is defined by (**??**) and $\mathcal{A}^*$ is defined by (1.12). Then we have*

$$\nabla f(x^*) = \sum_{i \in \mathcal{A}^*} \lambda_i a_i, \tag{1.14}$$

for some nonnegative coefficients λ_i, $i \in \mathcal{A}^*$.

Properly speaking, the KKT conditions are defined as the conditions defining feasibility and the active set $\mathcal{A}^*$, along with the condition (1.14). A succinct way to state them is as follows:

$$0 \leq \lambda \perp Ax^* - b \geq 0, \tag{1.15a}$$

$$\nabla f(x^*) = \sum_{j=1}^{m} \lambda_j a_j = A^T \lambda, \tag{1.15b}$$

where A is the matrix whose rows are $a_1^T, a_2^T, \ldots$, and $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)^T$. Note that the first condition (1.15a) guarantees feasibility of x^*. It also ensures that the components of λ that correspond to *inactive* constraints are set to zero, because $a_j^T x^* - b_j > 0 \Rightarrow \lambda_j = 0$ by the perpendicularity operator. Thus, the inactive terms contribute nothing to the sum in (1.15b), so the right-hand sides of (1.15b) and (1.14) are the same.

Example. We illustrate the KKT conditions with a simple example in two dimensions:

$$\min\, (x_1+1)^2 + 2(x_2+1)^2 \ \text{ subject to } \ x_2 \geq 0,\ x_1 + x_2 \geq 0. \tag{1.16}$$

We can express the constraints in the form (1.11) by defining:

$$m = 2, \quad a_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \; a_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad b_1 = 0, \; b_2 = 0.$$

The problem is shown in Figure 1.2, and we can see that the solution is at $x^* = (0,0)^T$. Let us check the KKT conditions at this point. Both constraints are active, that is, $\mathcal{A}^* = \{1,2\}$. We have

$$\nabla f(x) = \begin{bmatrix} 2(x_1+1) \\ 4(x_2+1) \end{bmatrix}, \quad \text{and so } \nabla f(x^*) = \nabla f(0) = \begin{bmatrix} 2 \\ 4 \end{bmatrix}.$$

We now seek $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ such that $\nabla f(x^*) = \lambda_1 a_1 + \lambda_2 a_2$, that is,

$$\begin{bmatrix} 2 \\ 4 \end{bmatrix} = \lambda_1 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \lambda_2 \\ \lambda_1 + \lambda_2 \end{bmatrix}.$$

It is easy to see that these conditions are satisfied (uniquely) by $\lambda_1 = 2$, $\lambda_2 = 2$, to the KKT conditions hold.

The *Lagrangian* provides a convenient way to state the KKT optimality conditions. It is a linear combination of the objective function and constraines defined as follows:

$$\mathcal{L}(x,\lambda) = f(x) - \sum_{i=1}^{m} \lambda_i (a_i^T x - b_i) = f(x) - \lambda^T (Ax - b). \tag{1.17}$$

The condition (1.15b) can be stated as

$$\nabla_x \mathcal{L}(x,\lambda) = \nabla f(x) - A^T \lambda = 0. \tag{1.18}$$

Therefore we can write the KKT conditions, including feasibility conditions, as follows:

$$0 \leq \lambda \perp Ax^* - b \geq 0, \quad \nabla_x \mathcal{L}(x^*, \lambda^*) = 0. \tag{1.19}$$