Sidework #2
RELEASE DATE: 11/26/2010
DUE DATE: NONE

1.1 KKT as Necessary Condition for Convex QP

Consider a quadratic programming problem

$$(P) \qquad \min_{\mathbf{w}} E(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T \mathrm{A}\mathbf{w} + \mathbf{b}^T\mathbf{w}$$
$$\text{subject to } \mathbf{p}_m^T\mathbf{w} \ge q_m \text{ for } m = 1, 2, \cdots, M.$$

with A being positive semi-definite. That is, (P) is convex. We will first prove the following remarkable fact of Karash-Kuhn-Tucker (KKT) conditions.

There exists $\boldsymbol{\alpha}^* \in R^M$ such that $(\mathbf{w}^*, \boldsymbol{\alpha}^*)$ satisfies the KKT conditions
if
(P) attains an optimal solution at $\mathbf{w}^*$.

The KKT conditions contain four parts:

$$\text{(K1: stationarity)} \qquad \nabla E(\mathbf{w}^*) = \mathrm{A}\mathbf{w}^* + \mathbf{b} = \sum_{m=1}^{M} \alpha_m^* \mathbf{p}_m.$$
$$\text{(K2: primal feasibility)} \qquad \mathbf{p}_m^T\mathbf{w}^* \ge q_m \text{ for } m = 1, 2, \cdots, M.$$
$$\text{(K3: dual feasibility)} \qquad \alpha_m^* \ge 0 \text{ for } m = 1, 2, \cdots, M.$$
$$\text{(K4: complementary slackness)} \qquad \alpha_m^*(\mathbf{p}_m^T\mathbf{w}^* - q_m) = 0 \text{ for } m = 1, 2, \cdots, M.$$

The proof that you will write below contains all the essential steps, but are not as rigorously written as the usual math texts.

(1) (preliminary) Show that for any $\mathbf{w} = \mathbf{w}^* + \eta\mathbf{v}$,

$$E(\mathbf{w}) = E(\mathbf{w}^*) + \eta\mathbf{v}^T\nabla E(\mathbf{w}^*) + \frac{1}{2}\eta^2\mathbf{v}^T\mathrm{A}\mathbf{v}.$$

(2) (stationarity, "if") Assume that (P) attains an optimal solution at $\mathbf{w}^*$ but

$$\nabla E(\mathbf{w}^*) = \sum_{m=1}^{M} \alpha_m^*\mathbf{p}_m + \mathbf{v}$$

with a nonzero $\mathbf{v}$ that is orthogonal to $\mathbf{p}_m$ for all m.

Let $\mathbf{w} = \mathbf{w}^* - \eta\mathbf{v}$ with $\eta > 0$. Use the fact in (preliminary) to show that when η is small enough, $E(\mathbf{w}) < E(\mathbf{w}^*)$ and $\mathbf{p}_m^T\mathbf{w} \ge q_m$ for all m. That is, $E(\mathbf{w})$ is a better solution than $E(\mathbf{w}^*)$.

Argue that you have proved that if (P) attains an optimal solution at $\mathbf{w}^*$, $(K1)$ must be satisfied.

(3) (dual feasibility, linearly independent constraints, "if") Assume that (P) attains an optimal solution at $\mathbf{w}^*$. We now know that

$$\nabla E(\mathbf{w}^*) = \sum_{m=1}^{M} \alpha_m^*\mathbf{p}_m$$

for some $\boldsymbol{\alpha}^*$. We shall first consider the case when the vectors $\mathbf{p}_m$ are linearly independent. Assume that dual feasibility is not satisfied. That is, without loss of generality, let $\alpha_1^* < 0$. Since $\mathbf{p}_m$ are linearly independent,

$$\mathbf{p}_1 = \sum_{m=2}^{M} \beta_m\mathbf{p}_m + \mathbf{v}$$

with a nonzero $\mathbf{v}$ that is orthogonal to $\mathbf{p}_m$ for $m = 2, 3, \cdots, M$.

Let $\mathbf{w} = \mathbf{w}^* + \eta\mathbf{v}$ with $\eta > 0$. Use the fact in (preliminary) to show that when η is small enough, $E(\mathbf{w}) < E(\mathbf{w}^*)$ and $\mathbf{p}_m^T\mathbf{w} \geq q_m$ for all m. That is, $E(\mathbf{w})$ is a better solution than $E(\mathbf{w}^*)$.

Argue that you have proved that if (P) attains an optimal solution at $\mathbf{w}^*$ and $\mathbf{p}_m$ are linearly independent, $(K3)$ must be satisfied.

(4) (dual feasibility, general constraints, "if") Assume that (P) attains an optimal solution at $\mathbf{w}^*$. If $\{\mathbf{p}_m\}_{m=1}^M$ are linearly dependent, but $\{\mathbf{p}_m\}_{m=2}^M$ are linearly independent. Assume that $(K1)$ is satisfied with some $\boldsymbol{\alpha}$ such that $\alpha_1 < 0$. Then, argue that there always exists an $\boldsymbol{\alpha}^*$ that satisfies both $(K1)$ and $(K3)$ with

$$\begin{aligned} \alpha_1^* &= 0 \\ \alpha_m^* &\geq 0 \text{ for } m = 2, 3, \cdots, M. \end{aligned}$$

Argue that you have proved that if (P) attains an optimal solution at $\mathbf{w}^*$, $(K3)$ must be satisfied.

(5) (complementary slackness, linearly independent constraints, "if") Assume that (P) attains an optimal solution at $\mathbf{w}^*$. We now know that

$$\nabla E(\mathbf{w}^*) = \sum_{m=1}^{M} \alpha_m^* \mathbf{p}_m$$

for some $\boldsymbol{\alpha}^*$ with non-negative components. We shall first consider the case when the vectors $\mathbf{p}_m$ are linearly independent. Assume that complementary slackness is not satisfied. That is, without loss of generality,

$$\alpha_1^*(\mathbf{p}_1^T\mathbf{w}^* - q_1) \neq 0.$$

From $(K2)$ and $(K3)$, it must mean that both $\alpha_1^* > 0$ and $\mathbf{p}_1^T\mathbf{w}^* > q_1$. Since $\mathbf{p}_m$ are linearly independent,

$$\mathbf{p}_1 = \sum_{m=2}^{M} \beta_m \mathbf{p}_m + \mathbf{v}$$

with a nonzero $\mathbf{v}$ that is orthogonal to $\mathbf{p}_m$ for $m = 2, 3, \cdots, M$.

Let $\mathbf{w} = \mathbf{w}^* - \eta\mathbf{v}$ with $\eta > 0$. Use the fact in (preliminary) to show that when η is small enough, $E(\mathbf{w}) < E(\mathbf{w}^*)$ and $\mathbf{p}_m^T\mathbf{w} \geq q_m$ for all m. That is, $E(\mathbf{w})$ is a better solution than $E(\mathbf{w}^*)$.

Argue that you have proved that if (P) attains an optimal solution at $\mathbf{w}^*$ and $\mathbf{p}_m$ are linearly independent, $(K4)$ must be satisfied.

(6) (complementary slackness, general constraints, "if") Use the same trick in (4) to argue that if (P) attains an optimal solution at $\mathbf{w}^*$ for any general linear constraints, $(K4)$ must be satisfied.

1.2 KKT as Sufficient Condition for Convex QP

We will now prove the sufficiency.

There exists $\boldsymbol{\alpha}^* \in R^M$ such that $(\mathbf{w}^*, \boldsymbol{\alpha}^*)$ satisfies the KKT conditions
only if
(P) attains an optimal solution at $\mathbf{w}^*$.

Assume that $(\mathbf{w}^*, \boldsymbol{\alpha}^*)$ satisfies the KKT conditions but $\mathbf{w}^*$ is not an optimal solution of (P). Then, there exists $\mathbf{w}$ that satisfies $(K2)$ with $E(\mathbf{w}) < E(\mathbf{w}^*)$.

(1) (convexity) Let $\mathbf{v} = \mathbf{w} - \mathbf{w}^*$ and consider $\mathbf{u} = \mathbf{w}^* + \eta\mathbf{v}$. Using the fact that $E(\mathbf{w})$ is convex, argue that $E(\mathbf{u}) < E(\mathbf{w}^*)$ for all $0 < \eta \leq 1$.

(2) (forbidden directions) Assume that $\mathbf{p}_m^T\mathbf{v} \geq 0$ for all m with $\alpha_m^* > 0$. Using the fact that $(\mathbf{w}^*, \boldsymbol{\alpha}^*)$ satisfies $(K1)$ and $(K3)$, argue that $\eta\mathbf{v}^T\nabla E(\mathbf{w}^*) \geq 0$. Then, show that $E(\mathbf{u}) \geq E(\mathbf{w}^*)$ for all η.

(3) (sufficiency) Argue from (1) and (2) that without loss of generality, there exists $\mathbf{p}_1^T\mathbf{v} < 0$ with $\alpha_1^* > 0$. Using the fact that $\mathbf{v} = \mathbf{w} - \mathbf{w}^*$, argue that $\mathbf{p}_1^T\mathbf{w}^* > \mathbf{p}_1^T\mathbf{w} \geq q_1$.

Argue that there is a violation of $(K4)$ and hence some assumptions must be wrong—namely, $\mathbf{w}^*$ should be an optimal solution!

Then, you can actually use KKT to prove the strong duality of convex QP problems (which includes the SVM problem we had in class). A standard treatment can be found on Pages 243 and 244 of the *Convex Optimization* textbook by Boyd and Vandenberghe (freely available online).