Lecture 10 - Linearly Constrained Problems: Separation $\rightarrow$ Alternative Theorems $\rightarrow$ Optimality Conditions

- A hyperplane

$$H = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T\mathbf{x} = b\} \ \ (\mathbf{a} \in \mathbb{R}^n \backslash \{\mathbf{0}\}, b \in \mathbb{R})$$

is said to strictly separate a point $\mathbf{y} \notin S$ from S if

$$\mathbf{a}^T\mathbf{y} > b$$

and

$$\mathbf{a}^T\mathbf{x} \leq b \text{ for all } \mathbf{y} \in S.$$

Theorem (separation of a point from a closed and convex set) Let $C \subseteq \mathbb{R}^n$ be a nonempty closed and convex set, and let $\mathbf{y} \notin C$. Then there exists $\mathbf{p} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ and $\alpha \in \mathbb{R}$ such that

$$\mathbf{p}^T\mathbf{y} > \alpha \text{ and } \mathbf{p}^T\mathbf{x} \leq \alpha \text{ for all } \mathbf{x} \in C.$$

Proof of the Separation Theorem

- By the second orthogonal projection theorem, the vector $\bar{\mathbf{x}} = P_C(\mathbf{y}) \in C$ satisfies
$$(\mathbf{y} - \bar{\mathbf{x}})^T(\mathbf{x} - \bar{\mathbf{x}}) \leq 0 \text{ for all } \mathbf{x} \in C,$$
which is the same as
$$(\mathbf{y} - \bar{\mathbf{x}})^T\mathbf{x} \leq (\mathbf{y} - \bar{\mathbf{x}})^T\bar{\mathbf{x}} \text{ for all } \mathbf{x} \in C.$$

Note p'x = alpha is the same as p' (x-xbar) = 0, which is the hyperplane going through xbar and orthgonal to p.
If the boundary of C is a smooth hypersurface, then that hyperplane is the tangent plane of the boundary surface at xbar.

- Denote $\mathbf{p} = \mathbf{y} - \bar{\mathbf{x}} \neq \mathbf{0}$ and $\alpha = (\mathbf{y} - \bar{\mathbf{x}})^T\bar{\mathbf{x}}$. Then
$$\mathbf{p}^T\mathbf{x} \leq \alpha \text{ for all } \mathbf{x} \in C$$

- On the other hand,
$$\mathbf{p}^T\mathbf{y} = (\mathbf{y} - \bar{\mathbf{x}})^T\mathbf{y} = (\mathbf{y} - \bar{\mathbf{x}})^T(\mathbf{y} - \bar{\mathbf{x}}) + (\mathbf{y} - \bar{\mathbf{x}})^T\bar{\mathbf{x}} = \|\mathbf{y} - \bar{\mathbf{x}}\|^2 + \alpha > \alpha.$$

Farkas Lemma - an Alternative Theorem

Farkas Lemma. Let $\mathbf{c} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{m\times n}$. Then exactly one of the following systems has a solution

I. $\mathbf{A}\mathbf{x} \le \mathbf{0}, \mathbf{c}^T\mathbf{x} > 0.$ intersection of a convex polyhedral cone and a half space in R^n

II. $\mathbf{A}^T\mathbf{y} = \mathbf{c}, \mathbf{y} \ge 0.$ intersection of a hyperplane and the positive orthant in R^m

Another equivalent formulation is the following.

Farkas Lemma - second Formulation Let $\mathbf{c} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{m\times n}$. Then the following two claims are equivalent:

(A) The implication $\mathbf{A}\mathbf{x} \le \mathbf{0} \Rightarrow \mathbf{c}^T\mathbf{x} \le 0$ holds true.

(B) There exists $\mathbf{y} \in \mathbb{R}^m_+$ such that $\mathbf{A}^T\mathbf{y} = \mathbf{c}$.

What does it mean?

Example. $\mathbf{A} = \begin{pmatrix} 1 & 5 \\ -1 & 2 \end{pmatrix}, \mathbf{c} = \begin{pmatrix} -1 \\ 9 \end{pmatrix},$

Claim: For A and c above, the set defined by I. is empty. Proof: Let y' = [1,2]. Then y' A = [-1 9] = c'. Since y>=0, this means if Ax<=0, then y'Ax <= 0, so c'x<=0. Q.E.D. This argument can easily to extended to show that if II is not empty, then I must be empty,or that, in the second formulation, (B)=>(A). The converse implication is less obvious, and the proof uses the separation theorem.

Proof of Farkas Lemma

- Suppose that system (B) is feasible: $\exists \mathbf{y} \in \mathbb{R}^m_+$ such that $\mathbf{A}^T\mathbf{y} = \mathbf{c}$.
- To see that the implication (A) holds, suppose that $\mathbf{A}\mathbf{x} \leq 0$ for some $\mathbf{x} \in \mathbb{R}^n$.
- Multiplying this inequality from the left by $\mathbf{y}^T$:

$$\mathbf{y}^T\mathbf{A}\mathbf{x} \leq 0.$$

- Hence,

$$\mathbf{c}^T\mathbf{x} \leq 0,$$

- Suppose that the implication (A) is satisfied, and let us show that the system (B) is feasible. Suppose in contradiction that system (B) is infeasible.
- Consider the following closed and convex (why?) set

$$S = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{A}^T\mathbf{y} \text{ for some } \mathbf{y} \in \mathbb{R}^m_+\}$$

- $\mathbf{c} \notin S$.

Proof Contd.

- By the separation theorem $\exists \mathbf{p} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ and $\alpha \in \mathbb{R}$ such that $\mathbf{p}^T\mathbf{c} > \alpha$ and

$$\mathbf{p}^T\mathbf{x} \leq \alpha \text{ for all } \mathbf{x} \in S. \tag{1}$$

- $\mathbf{0} \in S \Rightarrow \alpha \geq 0 \Rightarrow \mathbf{p}^T\mathbf{c} > 0.$
- (1) is equivalent to

$$\mathbf{p}^T\mathbf{A}^T\mathbf{y} \leq \alpha \text{ for all } \mathbf{y} \geq \mathbf{0}$$

or to

$$(\mathbf{A}\mathbf{p})^T\mathbf{y} \leq \alpha \text{ for all } \mathbf{y} \geq \mathbf{0}, \tag{2}$$

>=0

- Therefore, $\mathbf{A}\mathbf{p} \leq \mathbf{0}$.
- Contradiction to the assertion that implication (A) holds.

Gordan's Alternative Theorem

Theorem. Let $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then exactly one of the following two systems has a solution.

(A) $\mathbf{Ax} < \mathbf{0}$.

(B) $\mathbf{p} \neq 0, \mathbf{A}^T\mathbf{p} = \mathbf{0}, \mathbf{p} \geq \mathbf{0}$.

Proof.

- Suppose that system (A) has a solution.
- Assume in contradiction that (B) is feasible: $\exists \mathbf{p} \neq \mathbf{0}$ satisfying $\mathbf{A}^T\mathbf{p} = \mathbf{0}, \mathbf{p} \geq \mathbf{0}$.

>=0

<0

- Multiplying the equality $\mathbf{A}^T\mathbf{p} = \mathbf{0}$ from the left by $\mathbf{x}^T$ yields $(\mathbf{Ax})^T\mathbf{p} = 0$, which is an impossible equality.
- Suppose that system (A) does not have a solution.
- System (A) is equivalent to (s is a scalar) to $\mathbf{Ax} + s\mathbf{e} \leq \mathbf{0}, s > 0$.
- or to $\tilde{\mathbf{A}} \begin{pmatrix} \mathbf{x} \\ s \end{pmatrix} \leq \mathbf{0}, \mathbf{c}^T \begin{pmatrix} \mathbf{x} \\ s \end{pmatrix} > 0$, where $\tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{A} & \mathbf{e} \end{pmatrix}$ and $\mathbf{c} = \mathbf{e}_{n+1}$.
- The infeasibility of (A) is thus equivalent to the infeasibility of the system

$$\tilde{\mathbf{A}}\mathbf{w} \leq \mathbf{0}, \mathbf{c}^T\mathbf{w} > 0, \mathbf{w} \in \mathbb{R}^{n+1}.$$

Proof of Gordan Contd.

- By Farkas' lemma, $\exists \mathbf{z} \in \mathbb{R}^m_+$ such that

$$\begin{pmatrix} \mathbf{A}^T \\ \mathbf{e}^T \end{pmatrix} \mathbf{z} = \mathbf{c}$$

- $\Leftrightarrow \exists \mathbf{z} \in \mathbb{R}^m_+ : \mathbf{A}^T \mathbf{z} = \mathbf{0}, \mathbf{e}^T \mathbf{z} = 1.$
- $\Leftrightarrow \exists \mathbf{0} \neq \mathbf{z} \in \mathbb{R}^m_+ : \mathbf{A}^T \mathbf{z} = \mathbf{0}.$
- $\Rightarrow$ System (B) is feasible.

KKT Conditions for Linearly Constrained Problems

Theorem (KKT conditions for linearly constrained problems - necessary optimality conditions)

Consider the minimization problem

Ex: compare this result with the Lagrange multiplier theorem in calculus (which allows general nonlinear, but only *equality*, constraints)

$$(\mathrm{P}) \quad \begin{array}{ll} \min & f(\mathbf{x}), \\ \text{s.t.} & \mathbf{a}_i^T\mathbf{x} \le b_i, i = 1, 2, \ldots, m \end{array}$$

where f is continuously differentiable over $\mathbb{R}^n$, $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m \in \mathbb{R}^n, b_1, b_2, \ldots, b_m \in \mathbb{R}$ and let $\mathbf{x}^*$ be a local minimum point of (P). Then there exist $\underline{\lambda_1, \lambda_2, \ldots, \lambda_m \ge 0}$ such that

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \mathbf{a}_i = \mathbf{0}. \tag{3}$$

lambda_i > 0 & a_i'x*-b_i < 0 cannot happen together, for each i.

and

$$\lambda_i(\mathbf{a}_i^T\mathbf{x}^* - b_i) = 0, \quad i = 1, 2, \ldots, m. \tag{4}$$

Note: We see (3), but not (4), in the Lagrange multiplier theorem.

Proof of KKT Theorem

- $\mathbf{x}^*$ is a local minimum $\Rightarrow$ $\mathbf{x}^*$ is a stationary point.
- $\nabla f(\mathbf{x}^*)^T(\mathbf{x}-\mathbf{x}^*) \geq 0$ for every $\mathbf{x} \in \mathbb{R}^n$ satisfying $\mathbf{a}_i^T\mathbf{x} \leq b_i$ for any $i = 1, 2, \ldots, m$.
- Denote the set of *active* constraints by

$$I(\mathbf{x}^*) = \{i : \mathbf{a}_i^T\mathbf{x}^* = b_i\}.$$

- Making the change of variables $\mathbf{y} = \mathbf{x} - \mathbf{x}^*$, we have

$$\nabla f(\mathbf{x}^*)^T\mathbf{y} \geq 0 \text{ for any } \mathbf{y} \in \mathbb{R}^m \text{ satisfying } \mathbf{a}_i^T(\mathbf{y}+\mathbf{x}^*) \leq b_i, i = 1, 2, \ldots, m.$$

- or $\nabla f(\mathbf{x}^*)^T\mathbf{y} \geq 0$ for any $\mathbf{y}$ satisfying

$$\begin{array}{ll} \mathbf{a}_i^T\mathbf{y} \leq 0 & i \in I(\mathbf{x}^*), \\ \mathbf{a}_i^T\mathbf{y} \leq b_i - \mathbf{a}_i^T\mathbf{x}^* & i \notin I(\mathbf{x}^*). \end{array}$$

- The second set of inequalities can be removed, that is, we will prove that

$$\mathbf{a}_i^T\mathbf{y} \leq 0 \text{ for all } i \in I(\mathbf{x}^*) \Rightarrow \nabla f(\mathbf{x}^*)^T\mathbf{y} \geq 0.$$

(Same as: grad f(x*)' y >= 0 for any y satisfying just a_i' y <= 0 for all i in I(x*).)

Proof Contd.

- Suppose then that $\mathbf{y}$ satisfies $\mathbf{a}_i^T\mathbf{y} \le 0$ for all $i \in I(\mathbf{x}^*)$
- Since $b_i - \mathbf{a}_i^T\mathbf{x}^* > 0$ for all $i \notin I(\mathbf{x}^*)$, it follows that there exists a small enough $\alpha > 0$ for which $\mathbf{a}_i^T(\alpha\mathbf{y}) \le b_i - \mathbf{a}_i^T\mathbf{x}^*$.
- Thus, since in addition $\mathbf{a}_i^T(\alpha\mathbf{y}) \le 0$ for any $i \in I(\mathbf{x}^*)$, it follows by the stationarity condition that $\nabla f(\mathbf{x}^*)^T\mathbf{y} \ge 0$.
- We have shown $\mathbf{a}_i^T\mathbf{y} \le 0$ for all $i \in I(\mathbf{x}^*) \Rightarrow \nabla f(\mathbf{x}^*)^T\mathbf{y} \ge 0$.
- By Farkas' lemma $\exists \lambda_i \ge 0, i \in I(\mathbf{x}^*)$ such that

$$-\nabla f(\mathbf{x}^*) = \sum_{i \in I(\mathbf{x}^*)} \lambda_i \mathbf{a}_i.$$

- Defining $\underline{\lambda_i = 0 \text{ for all } i \notin I(\mathbf{x}^*)}$ we get that $\lambda_i(\mathbf{a}_i^T\mathbf{x}^* - b_i) = 0$ for all $i \in \{1, 2, \ldots, m\}$ and

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \mathbf{a}_i = \mathbf{0}.$$

The Convex Case

Theorem [KKT conditions for convex linearly constrained problems - necessary and sufficient optimality conditions]
Consider the minimization problem

$$
(\mathrm{P}) \qquad \begin{array}{ll} \min & f(\mathbf{x}), \\ \text{s.t.} & \mathbf{a}_i^T \mathbf{x} \le b_i, i = 1, 2, \ldots, m \end{array}
$$

where f is a convex continuously differentiable function over $\mathbb{R}^n$, $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m \in \mathbb{R}^n, b_1, b_2, \ldots, b_m \in \mathbb{R}$ and let $\mathbf{x}^*$ be a feasible solution of (P). Then $\mathbf{x}^*$ is an optimal solution if and only if there exist $\lambda_1, \lambda_2, \ldots, \lambda_m \ge 0$ such that

$$
\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \mathbf{a}_i = \mathbf{0}. \tag{5}
$$

and

$$
\lambda_i (\mathbf{a}_i^T \mathbf{x}^* - b_i) = 0, \quad i = 1, 2, \ldots, m. \tag{6}
$$

Proof of KKT in Convex Case

- Necessity was proven.
- Suppose that $\mathbf{x}^*$ is a feasible solution of (P) satisfying (5) and (6). Let $\mathbf{x}$ be a feasible solution of (P).
- Define the function

$$h(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i (\mathbf{a}_i^T \mathbf{x} - b_i).$$

1. It is convex, isn't it?
2. What is grad h(x) ?

(5)

- $\nabla h(\mathbf{x}^*) = \mathbf{0} \Rightarrow \mathbf{x}^*$ is a minimizer of h over $\mathbb{R}^n$.
-

$$f(\mathbf{x}^*) = f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i (\mathbf{a}_i^T \mathbf{x}^* - b_i) \leq f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i (\mathbf{a}_i^T \mathbf{x} - b_i) \leq f(\mathbf{x}),$$

x is feasible

Problems with Equality and Inequality Constraints

Theorem [KKT conditions for linearly constrained problems]
Consider the minimization problem

$$
\text{(Q)} \quad \begin{array}{ll} \min & f(\mathbf{x}), \\ \text{s.t.} & \mathbf{a}_i^T \mathbf{x} \le b_i, i = 1, 2, \ldots, m, \\ & \mathbf{c}_j^T \mathbf{x} = d_j, \quad j = 1, 2, \ldots, p. \end{array}
$$

Equiv. to: c_j' x <= d_j
& -c_j' x <= -d_j

where f cont. dif., $\mathbf{a}_i, \mathbf{c}_j \in \mathbb{R}^n$, $b_i, d_j \in \mathbb{R}$.

(i) **(necessity of the KKT conditions)** If $\mathbf{x}^*$ is a local minimum of (Q), then there exist $\underline{\lambda_1, \lambda_2, \ldots, \lambda_m \ge 0}$ and $\underline{\mu_1, \mu_2, \ldots, \mu_p \in \mathbb{R}}$ such that

$$
\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \mathbf{a}_i + \sum_{j=1}^p \mu_j \mathbf{c}_j = \mathbf{0}, \tag{7}
$$

Another easy trick: Any real number can be written as the difference of two non-negative numbers.

$$
\lambda_i(\mathbf{a}_i^T \mathbf{x}^* - b_i) = 0, \quad i = 1, 2, \ldots, m. \tag{8}
$$

(ii) **(sufficiency in the convex case)** If f is convex over $\mathbb{R}^n$ and $\mathbf{x}^*$ is a feasible solution of (Q) for which there exist $\lambda_1, \ldots, \lambda_m \ge 0$ and $\mu_1, \ldots, \mu_p \in \mathbb{R}$ such that (7) and (8) are satisfied, then $\mathbf{x}^*$ is an optimal solution of (Q).

Representation Via the Lagrangian

Given the a problem

$$\text{(NLP)} \quad \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & g_i(\mathbf{x}) \leq 0, i = 1, 2, \ldots, m, \\ & h_j(\mathbf{x}) = 0, j = 1, 2, \ldots, p. \end{array}$$

The associated Lagrangian function os

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j h_j(\mathbf{x}).$$

The KKT conditions can be written as

$$\begin{aligned} \nabla_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= \nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j \nabla h_j(\mathbf{x}^*) = \mathbf{0} \\ \lambda_i g_i(\mathbf{x}^*) &= 0, \quad i = 1, 2, \ldots, m. \end{aligned}$$

Examples

- $$\begin{array}{ll} \min & \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) \\ \text{s.t.} & x_1 + x_2 + x_3 = 3. \end{array}$$

- $$\begin{array}{ll} \min & x_1^2 + 2x_2^2 + 4x_1x_2 \\ \text{s.t.} & x_1 + x_2 = 1, \\ & x_1, x_2 \geq 0. \end{array}$$

In class

Projection onto Affine Spaces

Lemma. Let C be the affine space

$$C = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}\},$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Then

assume rank m

$$P_C(\mathbf{y}) = \mathbf{y} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{y} - \mathbf{b}).$$

Proof. In class

min ||x - y||^2 s.t. A x = b.

Everything is convex (in fact it is a QP), apply the KKT result. (In this case, it says that the usual Lagrange multiplier condition is both necessary and sufficient for optimality.)

(Of course, if you know your linear algebra, you can derive the same result without the KKT theorem, just like you don't need calculus to tell you that a x^2 + b x has its minimizer at x=-b/(2a).)

Orthogonal Projection onto Hyperplanes

Consider the hyperplane

$$H = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T\mathbf{x} = b\} \quad (\mathbf{0} \neq \mathbf{a} \in \mathbb{R}^n, b \in \mathbb{R}).$$

Then by the previous slide:

$$P_H(\mathbf{y}) = \mathbf{y} - \mathbf{a}(\mathbf{a}^T\mathbf{a})^{-1}(\mathbf{a}^T\mathbf{y} - b) = \mathbf{y} - \frac{\mathbf{a}^T\mathbf{y} - b}{\|\mathbf{a}\|^2}\mathbf{a}.$$

Lemma **(distance of a point from a hyperplane)** Let $H = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T\mathbf{x} = b\}$, where $\mathbf{0} \neq \mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Then

$$d(\mathbf{y}, H) = \frac{|\mathbf{a}^T\mathbf{y} - b|}{\|\mathbf{a}\|}.$$

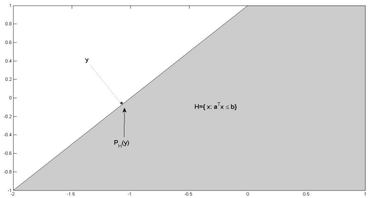
Proof.

$$d(\mathbf{y}, H) = \|\mathbf{y} - P_H(\mathbf{y})\| = \left\|\mathbf{y} - \left(\mathbf{y} - \frac{\mathbf{a}^T\mathbf{y} - b}{\|\mathbf{a}\|^2}\mathbf{a}\right)\right\| = \frac{|\mathbf{a}^T\mathbf{y} - b|}{\|\mathbf{a}\|}.$$

Orthogonal Projection onto Half-Spaces

Let $H^- = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T\mathbf{x} \leq b\}$, where $\mathbf{0} \neq \mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Then

$$P_{H^-}(\mathbf{x}) = \mathbf{x} - \frac{[\mathbf{a}^T\mathbf{x} - b]_+}{\|\mathbf{a}\|^2}\mathbf{a}$$



Proof. In class

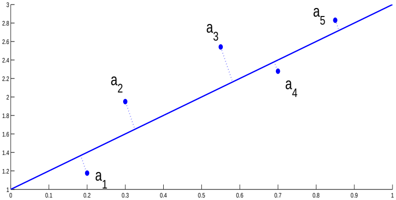
Orthogonal Regression

Q: Is this different from the standard least square problem (aka linear regression) in statistics? (See Chapter 3.)

A: Yes, quite different!

- $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$.
- For a given $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$, we define the hyperplane:

$$H_{\mathbf{x},y} := \left\{\mathbf{a} \in \mathbb{R}^n : \mathbf{x}^T\mathbf{a} = y\right\}.$$



- In the orthogonal regression problem we seek to find a nonzero vector $\mathbf{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$ such that the sum of squared Euclidean distances between the points $\mathbf{a}_1, \ldots, \mathbf{a}_m$ to $H_{\mathbf{x},y}$ is minimal:

$$\min_{\mathbf{x},y} \left\{\sum_{i=1}^{m} d(\mathbf{a}_i, H_{\mathbf{x},y})^2 : \mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n, y \in \mathbb{R}\right\}.$$

Orthogonal Regression

Not a convex problem, and this time we have no intention to convexify it !

- $d(\mathbf{a}_i, H_{\mathbf{x},y})^2 = \frac{(\mathbf{a}_i^T\mathbf{x}-y)^2}{\|\mathbf{x}\|^2}, \quad i = 1, \ldots, m.$
- The Orthogonal Regression problem is the same as

$$\min \left\{ \sum_{i=1}^{m} \frac{(\mathbf{a}_i^T\mathbf{x} - y)^2}{\|\mathbf{x}\|^2} : \mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n, y \in \mathbb{R} \right\}.$$

- Fixing $\mathbf{x}$ and minimizing first with respect to y we obtain that the optimal y is given by $y = \frac{1}{m}\sum_{i=1}^m \mathbf{a}_i^T\mathbf{x} = \frac{1}{m}\mathbf{e}^T\mathbf{A}\mathbf{x}$.
- Using the above expression for y we obtain that

$$\begin{aligned}
\sum_{i=1}^{m}\left(\mathbf{a}_i^T\mathbf{x} - y\right)^2 &= \sum_{i=1}^{m}\left(\mathbf{a}_i^T\mathbf{x} - \frac{1}{m}\mathbf{e}^T\mathbf{A}\mathbf{x}\right)^2 \\
&= \sum_{i=1}^{m}(\mathbf{a}_i^T\mathbf{x})^2 - \frac{2}{m}\sum_{i=1}^{m}(\mathbf{e}^T\mathbf{A}\mathbf{x})(\mathbf{a}_i^T\mathbf{x}) + \frac{1}{m}(\mathbf{e}^T\mathbf{A}\mathbf{x})^2 \\
&= \sum_{i=1}^{m}(\mathbf{a}_i^T\mathbf{x})^2 - \frac{1}{m}(\mathbf{e}^T\mathbf{A}\mathbf{x})^2 = \|\mathbf{A}\mathbf{x}\|^2 - \frac{1}{m}(\mathbf{e}^T\mathbf{A}\mathbf{x})^2 \\
&= \mathbf{x}^T\mathbf{A}^T\left(\mathbf{I}_m - \frac{1}{m}\mathbf{e}\mathbf{e}^T\right)\mathbf{A}\mathbf{x}.
\end{aligned}$$

Orthogonal Regression

- Therefore, a reformulation of the problem is

$$\min_{\mathbf{x}} \left\{ \frac{\mathbf{x}^T[\mathbf{A}^T(\mathbf{I}_m - \frac{1}{m}\mathbf{e}\mathbf{e}^T)\mathbf{A}]\mathbf{x}}{\|\mathbf{x}\|^2} : \mathbf{x} \neq \mathbf{0} \right\}.$$

Proposition. An optimal solution of the orthogonal regression problem $(\mathbf{x}, y)$ where $\mathbf{x}$ is an eigenvector of $\mathbf{A}^T(\mathbf{I}_m - \frac{1}{m}\mathbf{e}\mathbf{e}^T)\mathbf{A}$ associated with the minimum eigenvalue and $y = \frac{1}{m}\sum_{i=1}^m \mathbf{a}_i^T\mathbf{x}$. The optimal function value of the problem is $\lambda_{\min}\left[\mathbf{A}^T(\mathbf{I}_m - \frac{1}{m}\mathbf{e}\mathbf{e}^T)\mathbf{A}\right]$.

We turn the (non-convex) optimization problem into an eigen- problem !