Next, define the dual function $g\colon \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$ by maximizing over the primal variable $\vec{x} \in \mathbb{R}^n$:

$$g(\vec{\lambda}, \vec{\mu}) = \min_{\vec{x} \in \mathbb{R}^n} L(\vec{x}, \vec{\lambda}, \vec{\mu}) \tag{8.107}$$

$$= \begin{cases} \sum_{i=1}^m (\vec{\lambda}_i^\top \vec{y}_i - \mu_i z_i), & \text{if } \sum_{i=1}^m (A_i^\top \vec{\lambda}_i + \mu_i \vec{b}_i) = \vec{c}, \\ -\infty, & \text{otherwise.} \end{cases} \tag{8.108}$$

The last equality follows by noticing that the objective is linear in $\vec{x}$; if its coefficient $\vec{c} - \sum_{i=1}^m (A_i^\top \vec{\lambda}_i + \mu_i \vec{b}_i) = \vec{0}$, then the objective value is the sum $\sum_{i=1}^m (\vec{\lambda}_i^\top \vec{y}_i - \mu_i z_i)$ regardless of the value of $\vec{x}$, while if $\vec{c} - \sum_{i=1}^m (A_i^\top \vec{\lambda}_i + \mu_i \vec{b}_i) \neq \vec{0}$ then we can make the objective value as low as we want by picking $\vec{x}$ appropriately. For instance, let $K > 0$ be a large positive number; then

$$\vec{x} = -K \left(\vec{c} - \sum_{i=1}^m (A_i^\top \vec{\lambda}_i + \mu_i \vec{b}_i) \right) \implies L(\vec{x}, \vec{\lambda}, \vec{\mu}) = -K \left\| \vec{c} - \sum_{i=1}^m (A_i^\top \vec{\lambda}_i + \mu_i \vec{b}_i) \right\|_2^2 + \sum_{i=1}^m (\vec{\lambda}_i^\top \vec{y}_i - \mu_i z_i) \tag{8.109}$$

which we can drive down to $-\infty$ by increasing K to $+\infty$. Thus, the dual problem is given by:

$$\max_{\substack{\vec{\lambda} \in \mathbb{R}^d \\ \vec{\mu} \in \mathbb{R}^m}} \quad \sum_{i=1}^m (\vec{\lambda}_i^\top \vec{y}_i - \mu_i z_i) \tag{8.110}$$

$$\text{s.t.} \quad \sum_{i=1}^m (A_i^\top \vec{\lambda}_i + \mu_i \vec{b}_i) = \vec{c}, \tag{8.111}$$

$$-(\vec{\lambda}_i, \mu_i) \preceq_{K_i} 0, \qquad \forall i \in \{1, \ldots, m\}. \tag{8.112}$$

where the last line uses the fact that for each K_i its conic dual is itself, or in standard SOCP form by:

$$\max_{\substack{\vec{\lambda} \in \mathbb{R}^d \\ \vec{\mu} \in \mathbb{R}^m}} \quad \sum_{i=1}^m (\vec{\lambda}_i^\top \vec{y}_i - \mu_i z_i) \tag{8.113}$$

$$\text{s.t.} \quad \left\| \sum_{i=1}^m (A_i^\top \vec{\lambda}_i + \mu_i \vec{b}_i) - \vec{c} \right\|_2 \leq 0 \tag{8.114}$$

$$\left\| \vec{\lambda}_i \right\|_2 \leq \mu_i, \qquad \forall i \in \{1, \ldots, m\}. \tag{8.115}$$

□

Direct Proof. We re-iterate the SOCP to take the dual of:

$$p^\star = \min_{\vec{x} \in \mathbb{R}^n} \quad \vec{c}^\top \vec{x} \tag{8.116}$$

$$\text{s.t.} \quad \|A_i \vec{x} - \vec{y}_i\|_2 \leq \vec{b}_i^\top \vec{x} + z_i, \qquad \forall i \in \{1, \ldots, m\}. \tag{8.117}$$

We add some variables to simplify. Namely, we introduce $\vec{u}_i \in \mathbb{R}^{d_i}$ and $w_i \in \mathbb{R}$ for each $i \in \{1, \ldots, m\}$. For convenience, we define $\vec{u} \doteq (\vec{u}_1, \ldots, \vec{u}_m) \in \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_m} = \mathbb{R}^d$, where again $d = \sum_{i=1}^m d_i$, and also define $\vec{w} = (w_1, \ldots, w_m) \in \mathbb{R}^m$. With these definitions, the SOCP can be written as

$$p^\star = \min_{\substack{\vec{x} \in \mathbb{R}^n \\ \vec{u} \in \mathbb{R}^d \\ \vec{w} \in \mathbb{R}^m}} \quad \vec{c}^\top \vec{x} \tag{8.118}$$

$$\text{s.t.} \quad \|\vec{u}_i\|_2 \leq w_i, \qquad \forall i \in \{1, \ldots, m\} \tag{8.119}$$

$$\vec{u}_i = A_i \vec{x} - \vec{y}_i, \qquad \forall i \in \{1, \ldots, m\} \tag{8.120}$$

$$w_i = \vec{b}_i^\top \vec{x} + z_i, \qquad \forall i \in \{1, \ldots, m\}. \tag{8.121}$$

We can thus define a Lagrangian for this system, say with dual variables $\vec{\lambda} \in \mathbb{R}^m$, $\vec{\eta} \in \mathbb{R}^d$ (with $\vec{\eta}_i \in \mathbb{R}^{d_i}$ for each i) and $\vec{\nu} \in \mathbb{R}^m$. We have

$$L(\vec{x}, \vec{u}, \vec{w}, \vec{\lambda}, \vec{\eta}, \vec{\nu}) = \vec{c}^\top \vec{x} + \sum_{i=1}^m \lambda_i (\|\vec{u}_i\|_2 - w_i) + \sum_{i=1}^m \vec{\eta}_i^\top (\vec{u}_i - A_i \vec{x} + \vec{y}_i) + \sum_{i=1}^m \nu_i (w_i - \vec{b}_i^\top \vec{x} - z_i) \tag{8.122}$$

$$= \left(\vec{c} - \sum_{i=1}^m (A_i^\top \vec{\eta}_i + \nu_i \vec{b}_i) \right)^\top \vec{x} + \sum_{i=1}^n (\lambda_i \|\vec{u}_i\|_2 + \vec{\eta}_i^\top \vec{u}_i) + \sum_{i=1}^m (-\lambda_i + \nu_i) w_i + \sum_{i=1}^m (\vec{\eta}_i^\top \vec{y}_i - \nu_i z_i). \tag{8.123}$$

Now define the dual function $g \colon \mathbb{R}^m \times \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$ by minimizing over the primal variables $(\vec{x}, \vec{u}, \vec{w}) \in \mathbb{R}^n \times \mathbb{R}^d \times \mathbb{R}^m$:

$$g(\vec{\lambda}, \vec{\eta}, \vec{\nu}) = \min_{\substack{\vec{x} \in \mathbb{R}^n \\ \vec{u} \in \mathbb{R}^d \\ \vec{w} \in \mathbb{R}^m}} L(\vec{x}, \vec{u}, \vec{w}, \vec{\lambda}, \vec{\eta}, \vec{\nu}) \tag{8.124}$$

$$= \begin{cases} \sum_{i=1}^m (\vec{\eta}_i^\top \vec{y}_i - \nu_i z_i), & \text{if } \vec{c} = \sum_{i=1}^m (A_i^\top \vec{\eta}_i + \nu_i \vec{b}_i) \\ & \text{and } \|\vec{\eta}\|_2 \le \lambda_i \quad \forall i \in \{1, \ldots, m\} \\ & \text{and } \lambda_i = \nu_i \quad \forall i \in \{1, \ldots, m\} \\ -\infty, & \text{otherwise.} \end{cases} \tag{8.125}$$

The last equality looks complicated and a bit magical but we methodically justify it here.

(a) The Lagrangian is linear in $\vec{x}$, indeed having the form

$$L = \left(\vec{c} - \sum_{i=1}^m (A_i^\top \vec{\eta}_i + \nu_i \vec{b}_i) \right)^\top \vec{x} \; + \text{ other terms not involving } \vec{x}, \tag{8.126}$$

so unless the coefficient $\vec{c} - \sum_{i=1}^m (A_i^\top \vec{\eta}_i + \nu_i \vec{b}_i)$ is $\vec{0}$, then we can make the Lagrangian arbitrarily negative by varying $\vec{x}$ while keeping $\vec{u}$ and $\vec{w}$ fixed. For instance, let $K > 0$ be a large positive number. Then

$$\vec{x} = -K \left(\vec{c} - \sum_{i=1}^m (A_i^\top \vec{\eta}_i + \nu_i \vec{b}_i) \right) \implies L = -K \left\| \vec{c} - \sum_{i=1}^m (A_i^\top \vec{\eta}_i + \nu_i \vec{b}_i) \right\|_2^2 + \text{ other terms not involving } \vec{x} \tag{8.127}$$

which we can drive down to $-\infty$ by sending $K \to \infty$. On the other hand, if $\vec{c} - \sum_{i=1}^m (A_i^\top \vec{\eta}_i + \nu_i \vec{b}_i) = \vec{0}$, then the first term in the Lagrangian is $\vec{0}$ regardless of the value of $\vec{x}$.

Thus, we have proved

$$\min_{\vec{x} \in \mathbb{R}^n} L = \begin{cases} \sum_{i=1}^n (\lambda_i \|\vec{u}_i\|_2 + \vec{\eta}_i^\top \vec{u}_i) + \sum_{i=1}^m (-\lambda_i + \nu_i) w_i + \sum_{i=1}^m (\vec{\eta}_i^\top \vec{y}_i - \nu_i z_i), & \text{if } \vec{c} = \sum_{i=1}^m (A_i^\top \vec{\eta}_i + \nu_i \vec{b}_i) \\ -\infty, & \text{otherwise.} \end{cases} \tag{8.128}$$

(b) Suppose that $\vec{c} = \sum_{i=1}^m (A_i^\top \vec{\eta}_i + \nu_i \vec{b}_i)$. The Lagrangian has the form

$$\min_{\vec{x} \in \mathbb{R}^n} L = \sum_{i=1}^m (\lambda_i \|\vec{u}_i\|_2 + \vec{\eta}_i^\top \vec{u}_i) + \text{ other terms not involving } \vec{u}. \tag{8.129}$$

Towards minimizing this expression over $\vec{u}$, we aim to solve the problem

$$\min_{\vec{u}_i \in \mathbb{R}^{d_i}} (\lambda_i \|\vec{u}_i\|_2 + \vec{\eta}_i^\top \vec{u}_i), \tag{8.130}$$

and collect the results for each i at the end. At first glance, it may seem hard to imagine this term blowing up at all. Towards finding out a possible blow-up case, if any, we use Cauchy-Schwarz to try to make the sum as small as possible. In particular, by Cauchy-Schwarz we have

$$\lambda_i \|\vec{u}_i\|_2 + \vec{\eta}_i^\top \vec{u}_i \geq \lambda_i \|\vec{u}_i\|_2 - \|\vec{\eta}_i\|_2 \|\vec{u}_i\|_2 = (\lambda_i - \|\vec{\eta}_i\|_2) \|\vec{u}_i\|_2 \tag{8.131}$$

with equality when $\vec{u}_i$ points in the opposite direction as $\vec{\eta}_i$, that is, $\vec{u}_i = -K\vec{\eta}_i$ for some $K \geq 0$. With this value of $\vec{u}_i$ (for varying $K \to \infty$) we shall try to make the Lagrangian go to $-\infty$. Indeed, in this case, we have

$$\lambda_i \|\vec{u}_i\|_2 + \vec{\eta}_i^\top \vec{u}_i = K(\lambda_i - \|\vec{\eta}_i\|_2) \|\vec{\eta}_i\|_2 . \tag{8.132}$$

First suppose that $\|\vec{\eta}_i\|_2 = 0$. Since $\lambda_i \geq 0$ in the Lagrangian formulation, we must have $\lambda_i \geq \|\vec{\eta}_i\|_2$, as indicated in the original equality. The optimal $\vec{u}_i$ is $\vec{u}_i = -K\vec{\eta}_i = \vec{0}$ (independently of the value of K), at which point the term in the Lagrangian becomes 0. We now deal with the non-edge case, assuming that $\vec{\eta}_i \neq \vec{0}$.

Suppose that $\lambda_i - \|\vec{\eta}_i\|_2 < 0$. Then by sending $K \to \infty$ with this choice of $\vec{u}_i = -K\vec{\eta}_i$ we drive the Lagrangian to $-\infty$. On the other hand, if $\lambda_i - \|\vec{\eta}_i\|_2 \geq 0$, then the minimizing choice for K is $K = 0$, so that $\vec{u}_i = \vec{0}$, and the term in the Lagrangian becomes 0. Thus,

$$\min_{\vec{u}_i \in \mathbb{R}^{d_i}} (\lambda_i \|\vec{u}_i\|_2 + \vec{\eta}_i^\top \vec{u}_i) = \begin{cases} 0, & \text{if } \lambda_i \geq \|\vec{u}_i\|_2 \\ -\infty, & \text{otherwise.} \end{cases} \tag{8.133}$$

Applying this logic to each $i \in \{1, \ldots, m\}$, we obtain

$$\min_{\substack{\vec{x} \in \mathbb{R}^n \\ \vec{u} \in \mathbb{R}^d}} L = \begin{cases} \sum_{i=1}^m (-\lambda_i + \nu_i) w_i + \sum_{i=1}^m (\vec{\eta}_i^\top \vec{y}_i - \nu_i z_i), & \text{if } \vec{c} = \sum_{i=1}^m (A_i^\top \vec{\eta}_i + \nu_i \vec{b_i}) \\ & \text{and } \lambda_i \geq \|\vec{u}_i\|_2 , \quad \forall i \in \{1, \ldots, m\} \\ -\infty, & \text{otherwise.} \end{cases} \tag{8.134}$$

(c) Suppose that $\vec{c} = \sum_{i=1}^m (A_i^\top \vec{\eta}_i + \nu_i \vec{b}_i)$ and $\lambda_i \geq \|\vec{u}_i\|_2$ for each i. Then the Lagrangian has the form

$$\min_{\substack{\vec{x} \in \mathbb{R}^n \\ \vec{u} \in \mathbb{R}^d}} L = \sum_{i=1}^m (\nu_i - \lambda_i) w_i + \text{ other terms not involving } \vec{w}. \tag{8.135}$$

Towards minimizing this expression over $\vec{w}$, we aim to solve the problem

$$\min_{w_i \in \mathbb{R}} (\nu_i - \lambda_i) w_i \tag{8.136}$$

and collect the results at the end. Thankfully this is much simpler than the rest of the calculations, since the objective is an unconstrained minimization of a linear function of a scalar w_i. If the coefficient $\nu_i - \lambda_i$ is nonzero, then we can thus blow up the objective in any direction by choosing w_i accordingly. Namely, if $\nu_i \neq \lambda_i$ then the choice of $w_i = -K(\nu_i - \lambda_i)$ for some positive scalar $K > 0$, simplifies the objective as $-K(\nu_i - \lambda_i)^2$. Since $(\nu_i - \lambda_i)^2 > 0$, taking $K \to \infty$ shows that the optimal value of the objective is $-\infty$. On the other hand, if $\nu_i = \lambda_i$ then the objective has value 0 independent of the choice of w_i. We have shown that

$$\min_{w_i \in \mathbb{R}} (\nu_i - \lambda_i) w_i = \begin{cases} 0, & \text{if } \nu_i = \lambda_i \\ -\infty, & \text{otherwise.} \end{cases} \tag{8.137}$$

Applying this logic to all $i \in \{1, \ldots, m\}$, we obtain

$$\min_{\substack{\vec{x}\in\mathbb{R}^n \\ \vec{u}\in\mathbb{R}^d \\ \vec{w}\in\mathbb{R}^m}} L = \begin{cases} \sum_{i=1}^m (\vec{\eta}_i^\top \vec{y}_i - \nu_i z_i), & \text{if } \vec{c} = \sum_{i=1}^m (A_i^\top \vec{\eta}_i + \nu_i \vec{b}_i) \\ & \text{and } \lambda_i \geq \|\vec{u}_i\|_2, \quad \forall i \in \{1, \ldots, m\} \\ & \text{and } \nu_i = \lambda_i, \quad \forall i \in \{1, \ldots, m\} \\ -\infty, & \text{otherwise.} \end{cases} \tag{8.138}$$

Now we can write down the dual problem as

$$d^\star = \max_{\substack{\vec{\lambda}\in\mathbb{R}_+^m \\ \vec{\eta}\in\mathbb{R}^d \\ \vec{\nu}\in\mathbb{R}^m}} g(\vec{\lambda}, \vec{\eta}, \vec{\nu}) \tag{8.139}$$

which simplifies to

$$d^\star = \max_{\substack{\vec{\lambda}\in\mathbb{R}^m \\ \vec{\eta}\in\mathbb{R}^d \\ \vec{\nu}\in\mathbb{R}^m}} \quad \sum_{i=1}^m (\vec{\eta}_i^\top \vec{y}_i - \nu_i z_i) \tag{8.140}$$

$$\text{s.t.} \quad \vec{c} = \sum_{i=1}^m (A_i^\top \vec{\eta}_i + \nu_i \vec{b}_i) \tag{8.141}$$

$$\lambda_i \geq \|\vec{u}_i\|_2, \qquad \forall i \in \{1, \ldots, m\} \tag{8.142}$$

$$\lambda_i = \nu_i, \qquad \forall i \in \{1, \ldots, m\} \tag{8.143}$$

$$\lambda_i \geq 0, \qquad \forall i \in \{1, \ldots, m\}. \tag{8.144}$$

Note that the constraint $\lambda_i \geq \|\vec{u}_i\|_2$ already implies $\lambda_i \geq 0$ since $\|\vec{u}_i\|_2 \geq 0$. Thus, we can rewrite the problem again as

$$d^\star = \max_{\substack{\vec{\lambda}\in\mathbb{R}^m \\ \vec{\eta}\in\mathbb{R}^d \\ \vec{\nu}\in\mathbb{R}^m}} \quad \sum_{i=1}^m (\vec{\eta}_i^\top \vec{y}_i - \nu_i z_i) \tag{8.145}$$

$$\text{s.t.} \quad \vec{c} = \sum_{i=1}^m (A_i^\top \vec{\eta}_i + \nu_i \vec{b}_i) \tag{8.146}$$

$$\lambda_i \geq \|\vec{u}_i\|_2, \qquad \forall i \in \{1, \ldots, m\} \tag{8.147}$$

$$\lambda_i = \nu_i, \qquad \forall i \in \{1, \ldots, m\}. \tag{8.148}$$

Now note that the last constraint forces $\vec{\lambda} = \vec{\nu}$. Thus, we can eliminate one of them; we choose arbitrarily to eliminate $\vec{\nu}$ by replacing it everywhere with $\vec{\lambda}$. This gives the dual problem as

$$d^\star = \max_{\substack{\vec{\lambda}\in\mathbb{R}^m \\ \vec{\eta}\in\mathbb{R}^d}} \quad \sum_{i=1}^m (\vec{\eta}_i^\top \vec{y}_i - \lambda_i z_i) \tag{8.149}$$

$$\text{s.t.} \quad \vec{c} = \sum_{i=1}^m (A_i^\top \vec{\eta}_i + \lambda_i \vec{b}_i) \tag{8.150}$$

$$\lambda_i \geq \|\vec{u}_i\|_2, \qquad \forall i \in \{1, \ldots, m\}. \tag{8.151}$$

To write this in SOCP form, we can write the affine constraint as a norm, obtaining

$$d^\star = \max_{\substack{\vec{\lambda}\in\mathbb{R}^m \\ \vec{\eta}\in\mathbb{R}^d}} \quad \sum_{i=1}^m (\vec{\eta}_i^\top \vec{y}_i - \lambda_i z_i) \tag{8.152}$$

$$\text{s.t.} \quad \left\| \vec{c} - \sum_{i=1}^{m} (A_i^\top \vec{\eta}_i + \lambda_i \vec{b}_i) \right\|_2 \leq 0 \tag{8.153}$$

$$\|\vec{u}_i\|_2 \leq \lambda_i, \qquad \forall i \in \{1, \ldots, m\}. \tag{8.154}$$

Thus, we have obtained that the dual of an SOCP is another SOCP. □

The following content is optional/out of scope for this semester. Regardless, it may be helpful to read it to gain context, or get a deeper understanding of various results.

8.5 (OPTIONAL) Semidefinite Programming

This section introduces the semidefinite program (SDP), one of the broadest classes of named optimization problems. We begin with its two forms, the inequality form and standard form, and their properties.

In this section, we will use $\succeq$ and $\preceq$ to denote inequalities between symmetric matrices. These are instances of generalized inequalities, as in Definition 176, associated with the (proper) cone of positive semidefinite matrices. All you need to know is: if we write $A \succeq 0$, this means A is symmetric positive semidefinite, whereas if $A \succeq B$ then it means $A - B$ is symmetric positive semidefinite. On the other hand, if we write $A \preceq 0$, this means $-A$ is symmetric positive semidefinite; that is, A is *symmetric negative semidefinite*, with all non-positive eigenvalues.

Recall that $\mathbb{S}^n$ is the set of $n \times n$ symmetric matrices, and $\mathbb{S}^n_+$ is the set of $n \times n$ symmetric positive semidefinite matrices.

Definition 201 (Semidefinite Program in Inequality Form)

A *semidefinite program in inequality form* is an optimization problem of the following form:

$$\min_{\vec{x} \in \mathbb{R}^n} \quad \vec{c}^\top \vec{x} \tag{8.155}$$

$$\text{s.t.} \quad F_0 + \sum_{i=1}^{n} x_i F_i \preceq 0, \tag{8.156}$$

where $\vec{c} \in \mathbb{R}^n$, and $F_0, F_1 \ldots, F_n \in \mathbb{S}^n$.

The expression $F_0 + \sum_{i=1}^n x_i F_i$ is referred to as a *linear matrix inequality*. The constraint set, i.e., the set of $\vec{x} \in \mathbb{R}^n$ such that $F_0 + \sum_{i=1}^n x_i F_i \preceq 0$, is called a *spectrahedron*.

Notice that we only require one linear matrix inequality in the definition. What if we had multiple? Suppose that we actually wanted to solve the problem

$$\min_{\vec{x} \in \mathbb{R}^n} \quad \vec{c}^\top \vec{x} \tag{8.157}$$

$$\text{s.t.} \quad F_0^{(1)} + \sum_{i=1}^{n} x_i F_i^{(1)} \preceq 0, \tag{8.158}$$

$$F_0^{(2)} + \sum_{i=1}^{n} x_i F_i^{(2)} \preceq 0, \tag{8.159}$$

$$\vdots \tag{8.160}$$

$$F_0^{(k)} + \sum_{i=1}^{n} x_i F_i^{(k)} \preceq 0. \tag{8.161}$$

This could be phrased using a single linear matrix inequality, and the problem would be

$$\min_{\vec{x}\in\mathbb{R}^n} \quad \vec{c}^\top \vec{x} \tag{8.162}$$

$$\text{s.t.} \quad \begin{bmatrix} F_0^{(1)} & & \\ & \ddots & \\ & & F_0^{(k)} \end{bmatrix} + \sum_{i=1}^n x_i \begin{bmatrix} F_i^{(1)} & & \\ & \ddots & \\ & & F_i^{(k)} \end{bmatrix} \preceq 0. \tag{8.163}$$

(If this reduction isn't clear to you, it's totally fine; try to prove it as an exercise.)

We now introduce another major standard form of SDPs.

Definition 202 (Semidefinite Program in Standard Form)

A *semidefinite program in standard form* is an optimization problem of the following form:

$$\min_{X\in\mathbb{S}^n} \quad \operatorname{tr}(CX) \tag{8.164}$$

$$\text{s.t.} \quad \operatorname{tr}(A_k X) = b_k, \qquad \forall k \in \{1,\dots,m\} \tag{8.165}$$

$$X \succeq 0, \tag{8.166}$$

where $C, A_1, \dots, A_m \in \mathbb{S}^n$, and $b_1, \dots, b_m \in \mathbb{R}$.

The first main theorem below establishes that the inequality and standard forms of an SDP are equivalent, in the sense that either can be reformulated as the other.

Theorem 203

An SDP in inequality form can be reformulated as an SDP in standard form, and vice versa.

Proof. Just for this proof, we introduce the notation $\operatorname{vec}\colon \mathbb{R}^{m\times n} \to \mathbb{R}^{mn}$, which takes an $m \times n$ matrix and unrolls it into an mn-length vector. With this notation, in fact, for two symmetric matrices $A, B \in \mathbb{S}^n$, we can write $\operatorname{tr}(AB) = \sum_{i=1}^n \sum_{j=1}^n A_{ij}B_{ij} = \operatorname{vec}(A)^\top \operatorname{vec}(B)$. On the other hand, we will sometimes need to access the element of $\operatorname{vec}(A)$ corresponding to A_{ij}; we denote this by $\operatorname{vec}(A)_{i,j}$ (where the comma makes it clear that the index is not the product of i and j). We will also use the notation $\operatorname{diag}\colon \mathbb{R}^n \to \mathbb{R}^{n\times n}$ which takes a vector and returns a diagonal matrix whose diagonal is the entries of this vector. This notation will greatly simplify things to follow.

"Inequality form $\implies$ Standard form": Let $\vec{c} \in \mathbb{R}^n$ and $F_0, F_1, \dots, F_n \in \mathbb{S}^d$, and consider the following SDP in inequality form:

$$\min_{\vec{x}\in\mathbb{R}^n} \quad \vec{c}^\top \vec{x} \tag{8.167}$$

$$\text{s.t.} \quad F_0 + \sum_{i=1}^n x_i F_i \preceq 0. \tag{8.168}$$

Our goal is to write it in the form

$$\min_{X\in\mathbb{S}^m} \quad \operatorname{tr}(CX) \tag{8.169}$$

$$\text{s.t.} \quad \operatorname{tr}(A_k X) = b_k, \qquad \forall k \in \{1,\dots,p\}, \tag{8.170}$$

$$X \succeq 0. \tag{8.171}$$

First, towards introducing the positive semidefinite constraint, we introduce a new variable $Y \in \mathbb{S}^d$, associated with $-(F_0 + \sum_{i=1}^n x_i F_i)$. That is, our original problem has the form

$$\min_{\substack{\vec{x}\in\mathbb{R}^n \\ Y\in\mathbb{S}^d}} \quad \vec{c}^\top \vec{x} \tag{8.172}$$

$$\text{s.t.} \quad Y + F_0 + \sum_{i=1}^n x_i F_i = 0, \tag{8.173}$$

$$Y \succeq 0. \tag{8.174}$$

Since we have a linear matrix equality, we can write it as a bunch of scalar equations to get it closer to the desired form, say in the following way:

$$\min_{\substack{\vec{x}\in\mathbb{R}^n \\ Y\in\mathbb{S}^d}} \quad \vec{c}^\top \vec{x} \tag{8.175}$$

$$\text{s.t.} \quad Y_{jk} + (F_0)_{jk} + \sum_{i=1}^n x_i (F_i)_{jk} = 0, \qquad \forall j,k \in \{1,\dots,d\} \tag{8.176}$$

$$Y \succeq 0. \tag{8.177}$$

But even this isn't quite right – after all, we require all decision variables to be encapsulated in a positive semidefinite matrix. The simplest way to do this is to form a block diagonal matrix where each block is an embedding of a decision variable into a positive semidefinite matrix; the large matrix will also be positive semidefinite in this case. Towards converting $\vec{x}$ to a positive semidefinite block, one could consider its diagonal matrix equivalent $\operatorname{diag}(\vec{x})$, but this would not be positive semidefinite unless all entries of $\vec{x}$ were positive. To ensure that this happens, we use slack variables, akin to the proof that general linear programs can be written in standard form.

Namely, associate vectors $\vec{x}^+, \vec{x}^- \in \mathbb{R}^n$ by the following formulae:

$$x_i^+ = \begin{cases} x_i, & x_i > 0 \\ 0, & x_i \le 0, \end{cases} \qquad x_i^- = \begin{cases} 0, & x_i > 0 \\ -x_i, & x_i < 0. \end{cases} \tag{8.178}$$

In this case $\vec{x} = \vec{x}^+ - \vec{x}^-$. Thus the original problem is equivalent to the reformulation

$$\min_{\substack{\vec{x}^+\in\mathbb{R}^n \\ \vec{x}^-\in\mathbb{R}^n \\ Y\in\mathbb{S}^n}} \quad \vec{c}^\top(\vec{x}^+ - \vec{x}^-) \tag{8.179}$$

$$\text{s.t.} \quad Y_{jk} + (F_0)_{jk} + \sum_{i=1}^n (x_i^+ - x_i^-)(F_i)_{jk} = 0, \qquad \forall j,k \in \{1,\dots,d\} \tag{8.180}$$

$$x_i^+ \ge 0, \qquad \forall i \in \{1,\dots,n\}, \tag{8.181}$$

$$x_i^- \ge 0, \qquad \forall i \in \{1,\dots,n\}, \tag{8.182}$$

$$Y \succeq 0. \tag{8.183}$$

Now we are in business; we can write all the inequality/definiteness constraints as

$$Z \doteq \begin{bmatrix} \operatorname{diag}(\vec{x}^+) & & \\ & \operatorname{diag}(\vec{x}^-) & \\ & & Y \end{bmatrix} \succeq 0. \tag{8.184}$$

This is the positive semidefiniteness constraint we want, so the decision variable is $Z \in \mathbb{S}^{2n+d}$. As notation, let $Z^{1,i} = \operatorname{diag}(\vec{x}^+)_{ii} = x_i^+$ be the i^{th} element of the first block, $Z^{2,i} = \operatorname{diag}(\vec{x}^-)_{ii} = x_i^-$ be the i^{th} element of the second block, and $Z^{3,ij} = Y_{ij}$ be the $(i,j)^{\text{th}}$ element of the third block. As notation for later, let $\mathcal{O}$ be the set of all indices in $\{1,\dots,2n+d\} \times \{1,\dots,2n+d\}$ which are not on the diagonal or part of the Y block, and thus must be set to zero; formally $\mathcal{O} = \{(i,j) \mid 1 \le i,j \le 2n+d, i \ne j, i \le 2n \text{ or } j \le 2n\}$.

Now, we have written our problem in the form

$$\min_{Z \in \mathbb{S}^{2n+d}} \quad \sum_{i=1}^{n} c_i (Z^{1,i} - Z^{2,i}) \tag{8.185}$$

$$\text{s.t.} \quad Z^{3,jk} + (F_0)_{jk} + \sum_{i=1}^{n} (Z^{1,i} - Z^{2,i})(F_i)_{jk} = 0, \qquad \forall j,k \in \{1,\dots,d\} \tag{8.186}$$

$$Z_{i,j} = 0, \qquad \forall (i,j) \in \mathcal{O}, \tag{8.187}$$

$$Z \succeq 0. \tag{8.188}$$

Notice that all constraints are affine or positive definite, and our objective is affine; by our discussion of affine functions, the affine constraints can be written in the form $\operatorname{tr}(A_k Z) = b_k$, and the objective can be written in the form $\operatorname{tr}(CZ)$, for some symmetric matrices A_k, C and scalars b_k, and $k \in \{1,\dots,m\}$ where $m = d^2 + |\mathcal{O}|$.[2] Thus we can write our problem as

$$\min_{Z \in \mathbb{S}^{2n+d}} \quad \operatorname{tr}(CZ) \tag{8.189}$$

$$\text{s.t.} \quad \operatorname{tr}(A_k Z) = b_k, \qquad \forall k \in \{1,\dots,m\} \tag{8.190}$$

$$Z \succeq 0, \tag{8.191}$$

as desired.

"Standard form $\implies$ Inequality form": Let $C, A_1, \dots, A_m \in \mathbb{S}^n$ be fixed symmetric matrices, and let $b_1, \dots, b_m \in \mathbb{R}$ be fixed scalars. Consider the following SDP in standard form:

$$\min_{X \in \mathbb{S}^n} \quad \operatorname{tr}(CX) \tag{8.192}$$

$$\text{s.t.} \quad \operatorname{tr}(A_k X) = b_k, \qquad \forall k \in \{1,\dots,m\} \tag{8.193}$$

$$X \succeq 0. \tag{8.194}$$

We want to write it in the form

$$\min_{\vec{x} \in \mathbb{R}^m} \quad \vec{c}^{\top} \vec{x} \tag{8.195}$$

$$\text{s.t.} \quad F_0 + \sum_{i=1}^{m} x_i F_i \preceq 0. \tag{8.196}$$

Notice by our notation that $\operatorname{tr}(CX) = \operatorname{vec}(C)^{\top} \operatorname{vec}(X)$ and that $\operatorname{tr}(A_k X) = \operatorname{vec}(A_k)^{\top} \operatorname{vec}(X)$. Thus, letting $\vec{x} \in \mathbb{R}^{n^2}$ be defined as $\vec{x} = \operatorname{vec}(X)$, our objective is linear in $\vec{x}$, since it is $\vec{c}^{\top}\vec{x}$ where $\vec{c} = \operatorname{vec}(C)$. Furthermore, our

[2] Careful readers may notice that the discussion on affine functions ensured something slightly different; namely, for an affine function f on symmetric matrices (or indeed all of $\mathbb{R}^{n \times n}$), there was some matrix A and scalar b such that $f(X) = \operatorname{tr}\left(A^{\top} X\right) + b$. In particular, the result did not guarantee that such an A could be symmetric. But certainly the matrix $(A + A^{\top})/2$ is symmetric, and for Z symmetric, we have $\operatorname{tr}\left(A^{\top} Z\right) = \operatorname{tr}\left([(A + A^{\top})/2]Z\right)$, so indeed, for an affine function $f \colon \mathbb{S}^n \to \mathbb{R}$ there exists some matrix $A \in \mathbb{S}^n$ and scalar $b \in \mathbb{R}$ such that $f(X) = \operatorname{tr}(AX) + b$.

equality constraints are affine, since they are $\vec{a}_k^\top \vec{x} = b_k$ where $\vec{a}_k = \mathrm{vec}(A_k)$. We express each equality constraint as a pair of linear matrix inequalities, since that is the only type of constraint we are permitted to have. Indeed, we have

$$\vec{a}_k^\top \vec{x} = b_k \iff \sum_{i=1}^m x_i (\vec{a}_k)_i = b_k \tag{8.197}$$

$$\iff -b_k + \sum_{i=1}^m x_i (\vec{a}_k)_i = 0 \tag{8.198}$$

$$\iff -b_k + \sum_{i=1}^m x_i (\vec{a}_k)_i \preceq 0 \text{ and } -\left(-b_k + \sum_{i=1}^m x_i (\vec{a}_k)_i\right) \preceq 0, \tag{8.199}$$

where we are using $\preceq$ for ordering on the space of 1×1 symmetric matrices, i.e., scalars. These are bona-fide linear matrix inequalities and will be combined with others, later, to form the full linear matrix inequality constraint for our problem.

The only constraint remaining that cannot easily be expressed in vectorized form is the constraint $X \succeq 0$. For this, we note that we are allowed to have a linear matrix inequality constraint, so we want to express $X \succeq 0$ in terms of a linear matrix inequality involving $\vec{x}$. This is difficult at first, so we handle it in the case $n = 2$ for an example. Write

$$X = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}, \qquad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}. \tag{8.200}$$

Notice that, since X is symmetric (and so $x_2 = x_3$), we can write X in terms of a linear combination of constant symmetric matrices, as follows

$$X = \begin{bmatrix} x_1 & x_2 \\ x_2 & x_4 \end{bmatrix} \tag{8.201}$$

$$= x_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{8.202}$$

$$= x_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \frac{1}{2} x_2 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \frac{1}{2} x_3 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{8.203}$$

$$= x_1 E^{11} + \frac{1}{2} x_2 (E^{12} + E^{21}) + \frac{1}{2} x_3 (E^{12} + E^{21}) + x_4 E^{22}, \tag{8.204}$$

where E^{ij} is defined as the $n \times n$ matrix with 1 in the $(i,j)^{\text{th}}$ coordinate and 0 elsewhere. Thus the positive semidefinite constraint can be replaced by the linear matrix inequality

$$X \succeq 0 \iff -\left(x_1 E^{11} + \frac{1}{2} x_2 (E^{12} + E^{21}) + \frac{1}{2} x_3 (E^{12} + E^{21}) + x_4 E^{22}\right) \preceq 0. \tag{8.205}$$

The general case goes the same way. We can say

$$X \succeq 0 \iff -\left(\sum_{i=1}^n x_{i,i} E^{ii} + \frac{1}{2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n x_{i,j} (E^{ij} + E^{ji})\right) \preceq 0 \tag{8.206}$$

where again $x_{i,j}$ refers to the element of $\vec{x}$ corresponding to the entry X_{ij}.

This gives a linear matrix inequality for the last constraint, and so all constraints can be represented by some linear matrix inequalities. Thus, by the discussion on reducing several linear matrix inequalities to a single one, all constraints

can be represented as a single linear matrix inequality of the form $F_0 + \sum_{i=1}^m x_i F_i$. Thus the original problem can be represented as

$$\min_{\vec{x} \in \mathbb{R}^m} \quad \vec{c}^\top \vec{x} \tag{8.207}$$

$$\text{s.t.} \quad F_0 + \sum_{i=1}^m x_i F_i \preceq 0, \tag{8.208}$$

where $m = n^2$, $\vec{c} = \operatorname{vec}(C)$, and the linear matrix inequality constraint is constructed in the aforementioned way. □

Theorem 204 (Dual of an SDP)
The dual of an SDP is an SDP.

Proof. Let $\vec{c} \in \mathbb{R}^n$, and let $F_0, F_1, \ldots, F_n \in \mathbb{S}^d$. Consider the following inequality-form SDP:

$$\min_{\vec{x} \in \mathbb{R}^n} \quad \vec{c}^\top \vec{x} \tag{8.209}$$

$$\text{s.t.} \quad F_0 + \sum_{i=1}^n x_i F_i \preceq 0. \tag{8.210}$$

We compute the conic dual of this problem. We know that the dual cone of $\mathbb{S}_+^d$ in $\mathbb{S}^d$ (equipped with the Frobenius inner product $\langle A,\ B \rangle_F = \operatorname{tr}(AB)$ and corresponding Frobenius norm) is simply $\mathbb{S}_+^d$ itself. Thus we can define the Lagrangian $L \colon \mathbb{R}^n \times \mathbb{S}_+^d$ as

$$L(\vec{x}, \Lambda) = \vec{c}^\top \vec{x} + \left\langle \Lambda,\ F_0 + \sum_{i=1}^n x_i F_i \right\rangle_F \tag{8.211}$$

$$= \vec{c}^\top \vec{x} + \operatorname{tr}\left(\Lambda \left(F_0 + \sum_{i=1}^n x_i F_i \right) \right) \tag{8.212}$$

$$= \vec{c}^\top \vec{x} + \operatorname{tr}(\Lambda F_0) + \sum_{i=1}^n x_i \operatorname{tr}(\Lambda F_i) \tag{8.213}$$

$$= \sum_{i=1}^n (c_i + \operatorname{tr}(\Lambda F_i)) x_i + \operatorname{tr}(\Lambda F_0). \tag{8.214}$$

Now, define the dual function $g \colon \mathbb{S}_+^n \to \mathbb{R}$ by minimizing over the primal variable $\vec{x}$:

$$g(\Lambda) = \min_{\vec{x} \in \mathbb{R}^d} L(\vec{x}, \Lambda) \tag{8.215}$$

$$= \min_{\vec{x} \in \mathbb{R}^d} \left(\sum_{i=1}^n (c_i + \operatorname{tr}(\Lambda F_i)) x_i + \operatorname{tr}(\Lambda F_0) \right) \tag{8.216}$$

$$= \operatorname{tr}(\Lambda F_0) + \sum_{i=1}^n \min_{x_i \in \mathbb{R}} (c_i + \operatorname{tr}(\Lambda F_i)) x_i \tag{8.217}$$

$$= \begin{cases} \operatorname{tr}(\Lambda F_0), & \text{if } \operatorname{tr}(\Lambda F_i) = -c_i, \quad \forall i \in \{1, \ldots, n\} \\ -\infty, & \text{otherwise.} \end{cases} \tag{8.218}$$

The last equality is because in each individual term $(c_i + \operatorname{tr}(\Lambda F_i)) x_i$, when minimizing over x_i, if $c_i + \operatorname{tr}(\Lambda F_i) \neq 0$ then we can always drive it to $-\infty$ by picking x_i to be large and of the opposite sign.

We can thus write the dual problem as

$$d^\star = \max_{\Lambda \in \mathbb{S}^d} \quad \operatorname{tr}(F_0 \Lambda) \tag{8.219}$$

$$\text{s.t.} \quad \operatorname{tr}(F_i \Lambda) = c_i, \qquad \forall i \in \{1, \ldots, n\}, \tag{8.220}$$

$$\Lambda \succeq 0, \tag{8.221}$$

which is an SDP in standard form. □

SDPs generalize all previously introduced classes of convex optimization problems: LPs, (convex) QPs, (convex) QCQPs, and SOCPs.

Theorem 205

SOCPs can be reformulated as SDPs.

Proof. We use the following useful characterization of second-order cone constraints as semidefinite constraints.

Claim. For $(\vec{x}, t) \in \mathbb{R}^{m+1}$, we have

$$\|\vec{x}\|_2 \le t \iff \begin{bmatrix} tI & \vec{x} \\ \vec{x}^\top & t \end{bmatrix} \succeq 0. \tag{8.222}$$

Proof of claim. We have

$$\begin{bmatrix} tI & \vec{x} \\ \vec{x}^\top & t \end{bmatrix} \succeq 0 \iff \begin{bmatrix} \vec{a} \\ b \end{bmatrix}^\top \begin{bmatrix} tI & \vec{x} \\ \vec{x}^\top & t \end{bmatrix} \begin{bmatrix} \vec{a} \\ b \end{bmatrix} \ge 0, \qquad \forall (\vec{a}, b) \in \mathbb{R}^{m+1} \tag{8.223}$$

$$\iff \begin{bmatrix} \vec{a} \\ b \end{bmatrix}^\top \begin{bmatrix} t\vec{a} + b\vec{x} \\ \vec{x}^\top \vec{a} + tb \end{bmatrix} \ge 0, \qquad \forall (\vec{a}, b) \in \mathbb{R}^{m+1} \tag{8.224}$$

$$\iff t\left(\|\vec{a}\|_2^2 + b^2\right) + 2b\vec{a}^\top \vec{x} \ge 0, \qquad \forall (\vec{a}, b) \in \mathbb{R}^{m+1}. \tag{8.225}$$

By Cauchy-Schwarz we have

$$t\left(\|\vec{a}\|_2^2 + b^2\right) + 2b\vec{a}^\top \vec{x} \ge t\left(\|\vec{a}\|_2^2 + b^2\right) - 2|b| \|\vec{a}\|_2 \|\vec{x}\|_2 \tag{8.226}$$

with equality when $\vec{a} = -K\vec{x}$ for some positive scalar $K > 0$. Thus

$$\begin{bmatrix} tI & \vec{x} \\ \vec{x}^\top & t \end{bmatrix} \succeq 0 \iff t\left(\|\vec{a}\|_2^2 + b^2\right) - 2|b| \|\vec{a}\|_2 \|\vec{x}\|_2 \ge 0, \qquad \forall (\vec{a}, b) \in \mathbb{R}^{m+1}. \tag{8.227}$$

Now by the AM-GM inequality (i.e., expanding the square on $(\|\vec{a}\|_2 - |b|)^2 \ge 0$), we have $\|\vec{a}\|_2^2 + b^2 \ge 2|b| \|\vec{a}\|_2$, with equality when $\|\vec{a}\|_2 = |b|$. This gives

$$t\left(\|\vec{a}\|_2^2 + b^2\right) - 2|b| \|\vec{a}\|_2 \|\vec{x}\|_2 \ge 2|b| \|\vec{a}\|_2 (t - \|\vec{x}\|_2) \tag{8.228}$$

with equality when $\|\vec{a}\|_2 = |b|$. Thus we have

$$\begin{bmatrix} tI & \vec{x} \\ \vec{x}^\top & t \end{bmatrix} \succeq 0 \iff 2|b| \|\vec{a}\|_2 (t - \|\vec{x}\|_2) \ge 0 \quad \forall (\vec{a}, b) \in \mathbb{R}^{m+1} \tag{8.229}$$

$$\iff t - \|\vec{x}\|_2 \ge 0 \tag{8.230}$$

as desired, so the claim is proved.

Now fix $\vec{c} \in \mathbb{R}^n$, fix $\vec{b}_1, \ldots, \vec{b}_m \in \mathbb{R}^n$, fix $A_1, \ldots, A_m$ so that $A_i \in \mathbb{R}^{d_i \times n}$, fix $\vec{y}_1, \ldots, \vec{y}_m$ so that $\vec{y}_i \in \mathbb{R}^{d_i}$, and fix $z_1, \ldots, z_m \in \mathbb{R}$. Consider the following generic SOCP:

$$\min_{\vec{x} \in \mathbb{R}^n} \quad \vec{c}^\top \vec{x} \tag{8.231}$$

$$\text{s.t.} \quad \|A_i \vec{x} - \vec{y}_i\|_2 \leq \vec{b}_i^\top \vec{x} + z_i, \qquad \forall i \in \{1, \ldots, m\}. \tag{8.232}$$

Let K^d be the second-order cone in $\mathbb{R}^d$. Notice that each cone constraint can be written in the form

$$\|A_i \vec{x} - \vec{y}_i\|_2 \leq \vec{b}_i^\top \vec{x} + z_i \tag{8.233}$$

$$\iff (A_i \vec{x} - \vec{y}_i, \vec{b}_i^\top \vec{x} + z_i) \in K^{d_i + 1} \tag{8.234}$$

$$\iff \begin{bmatrix} (\vec{b}_i^\top \vec{x} + z_i) I & A_i \vec{x} - \vec{y}_i \\ (A_i \vec{x} - \vec{y}_i)^\top & \vec{b}_i^\top \vec{x} + z_i \end{bmatrix} \succeq 0 \tag{8.235}$$

$$\iff \begin{bmatrix} z_i I & -\vec{y}_i \\ -\vec{y}_i^\top & z_i \end{bmatrix} + \sum_{j=1}^{n} x_j \begin{bmatrix} (\vec{b}_i)_j I & (A_i)_j \\ (A_i)_j^\top & (\vec{b}_i)_j \end{bmatrix} \succeq 0 \tag{8.236}$$

where $(\vec{b}_i)_j$ is the j^{th} entry of $\vec{b}_i$, and $(A_i)_j$ is the j^{th} column of A_i. Anyways, this is a linear matrix inequality (after some reshuffling of terms).

The conversion from SOCP to SDP consists of converting all second-order cone constraints to small linear matrix inequalities, then combining them to form one larger linear matrix inequality, which defines the constraint set of the inequality-form SDP. The objective function is already linear, so the resulting SDP is in the "standard" inequality form. Thus, we have reduced the original SOCP to an SDP. □

Note that in practice, this reduction is often extremely costly; SDPs are hard to solve at large scale, while SOCPs are much easier.

The above content is optional/out of scope for this semester, but now we resume the required/in scope content.

8.6 General Taxonomy

We conclude this chapter with a taxonomy of problems that we have discussed until now:

$$\texttt{LPs} \subset \texttt{Convex QPs} \subset \texttt{Convex QCQPs} \subset \texttt{SOCPs} \subset \texttt{SDPs} \subset \texttt{Convex Problems} \tag{8.237}$$

All inclusions are strict, i.e., none of the classes is equivalent to any of the others.

For extra optional reading, you may also look into *geometric programs* (GPs), which are nonconvex programs that can be turned into convex programs with a change of variables; and mixed-integer programs (MIPs), which are useful in practice to incorporate integer constraints, but difficult to solve exactly. All such material is out of scope of the course.

Chapter 9

Regularization and Sparsity

Relevant sections of the textbooks:

- [2] Chapters 9, 12, 13.

9.1 Recapping Ridge Regression and Defining LASSO

The first example of regularization we saw was ridge regression. In this section, we'll review ridge regression. The most basic perspective of ridge regression focuses on the additional term we add to the objective function. In ridge regression, we solve the following problem:

$$\min_{\vec{x}\in\mathbb{R}^n} \left\{ \|A\vec{x} - \vec{y}\|_2^2 + \lambda \|\vec{x}\|_2^2 \right\}. \tag{9.1}$$

This is different from the OLS problem due to the additional $\lambda \|\vec{x}\|_2^2$ term, which can be thought of as a *regularizer* (i.e., a penalty) for having large $\vec{x}$ values. The λ parameter controls the strength of the penalty and is usually called a *regularization parameter*. In this sense, ridge regression is *regularized least squares*. More generally, we may define regularization as follows.

Definition 206 (Regularization)
Consider the optimization problem

$$p^\star = \min_{\vec{x}\in\Omega} f_0(\vec{x}). \tag{9.2}$$

For a given function $R\colon \Omega \to \mathbb{R}_+$ (the *regularizer*) and a *regularization parameter* $\lambda > 0$, the *regularized version* of the above problem is the problem

$$p_\lambda^\star = \min_{\vec{x}\in\Omega}\{f_0(\vec{x}) + \lambda R(\vec{x})\}. \tag{9.3}$$

Here λ controls the strength of the regularization.

In general, the original problem and the regularized problem *do not have the same solutions*, nor do versions of the regularized problem with different λ parameter. One need only consider ridge regression to keep this in mind; for a fixed A and $\vec{y}$, increasing λ will decrease the norm of the solution to the ridge regression problem, and sending it to 0 (i.e., recovering unregularized least squares) will increase the norm of the solution.

One example of regularization is the ℓ^2-norm penalty $R(\vec{x}) = \|\vec{x}\|_2^2$, which (when combined with $f_0(\vec{x}) = \|A\vec{x} - \vec{y}\|_2^2$) yields ridge regression. Another example is the elastic-net regression, which we covered briefly as an

example when discussing convexity. But the main objective of this chapter is to look at the so-called LASSO regression problem, which uses an ℓ^1-norm regularizer. Recall that for a vector $\vec{x} \in \mathbb{R}^n$, its ℓ^1-norm is defined as $\|\vec{x}\|_1 = \sum_{i=1}^n |x_i|$.

Definition 207 (LASSO Regression)

Lett $A \in \mathbb{R}^{m \times n}$, $\vec{y} \in \mathbb{R}^m$, and $\lambda > 0$. The *LASSO regression* problem is:

$$\min_{\vec{x} \in \mathbb{R}^n} \left\{ \|A\vec{x} - \vec{y}\|_2^2 + \lambda \|\vec{x}\|_1 \right\}. \tag{9.4}$$

Here are some key properties of the LASSO regression problem.

Proposition 208

Consider the LASSO regression problem

$$\min_{\vec{x} \in \mathbb{R}^n} f_0(\vec{x}) \qquad \text{where} \qquad f_0(\vec{x}) \doteq \|A\vec{x} - \vec{y}\|_2^2 + \lambda \|\vec{x}\|_1 . \tag{9.5}$$

(a) The function $f_0 \colon \mathbb{R}^n \to \mathbb{R}$ is convex.

(b) If A has full column rank then f_0 is μ-strongly convex with $\mu = 2\sigma_n\{A\}^2$.

(c) A solution $\vec{x}^\star \in \operatorname{argmin}_{\vec{x} \in \mathbb{R}^n} f_0(\vec{x})$ always exists.

(d) If A has full column rank then the above solution is unique.

This picture is *very* different from ridge regression, where we are guaranteed that a solution always exists, is unique, and solvable in closed form. The question then becomes: why do we even care about the LASSO problem at all? The basic answer is that *it induces sparsity in the solution*, i.e., solutions to LASSO usually tend to have few nonzero entries. This sparsity is useful for applications in high-dimensional statistics and machine learning, as it reveals a certain structure — in words, it points out which "features" are the most relevant to the regression. In the following sections, we will observe how this sparsity emerges, both geometrically and algebraically.

9.2 Understanding the Difference Between the ℓ^2-Norm and the ℓ^1-Norm

In this section, we attempt to build more intuition about the difference between the ℓ^2-norm and the ℓ^1-norm. We do this by solving some problems which use the ℓ^2 norm, then replace it with the ℓ^1 norm and solve this new problem. Besides giving us intuition, it will help us learn how to analyze the LASSO problem.

Here is a diagram of the norm balls of the ℓ^1 (blue) and ℓ^2 (red) norms in $n = 1$ dimensions:

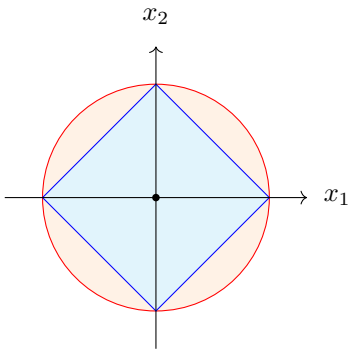

Figure 9.1: The ℓ^1 and ℓ^2 norm balls in $n = 2$ dimensions. Recall that the ℓ^p-norm ball is defined as the set of vectors $\vec{v}$ such that $\|\vec{v}\|_p \leq 1$.

The border of the norm balls are the points where each norm is equal to 1. Notice the difference in the geometry of these norm balls. The ℓ^2 norm ball is circular, while the ℓ^1 norm ball has distinctive corners.

In fact, these corners hint at a key difference between these norms: the ℓ^2 norm is differentiable everywhere, but $\|\vec{x}\|_1$ is *not differentiable* when *any* $x_i = 0$. These corners will help us understand how the ℓ^1 norm regularizer induces sparsity in the solution, and also inform our analysis of problems involving the ℓ^1-norm, including LASSO.

Example 209 (Least ℓ^1-Norm)**.** Recall that we solved the problem

$$\min_{\vec{x}\in\mathbb{R}^n} \quad \|\vec{x}\|_2^2 \tag{9.6}$$

$$\text{s.t.} \quad A\vec{x} = \vec{y}. \tag{9.7}$$

Using the KKT conditions, namely stationarity, we found an explicit solution to this problem: $\vec{x}^\star = A^\top(AA^\top)^{-1}\vec{y}$. Now let us replace the ℓ^2 norm with an ℓ^1 norm; we obtain the problem

$$\min_{\vec{x}\in\mathbb{R}^n} \quad \|\vec{x}\|_1 \tag{9.8}$$

$$\text{s.t.} \quad A\vec{x} = \vec{y}. \tag{9.9}$$

We cannot apply stationarity to this problem because the objective is non-differentiable. Thus, this problem seems intractable to solve by hand, at least for the moment. Instead, let us formulate it as a linear program. As before, we represent each x_i as the difference of non-negative numbers which sum to $|x_i|$. More formally, we introduce slack variables $\vec{x}^+, \vec{x}^- \in \mathbb{R}^n$ such that for each $i \in \{1, \ldots, n\}$ we have $x_i^+ \geq 0$, $x_i^- \geq 0$, $x_i^+ - x_i^- = x_i$, and $x_i^+ + x_i^- = |x_i|$. Thus we can rewrite the problem using the following linear program:

$$\min_{\vec{x}^+,\vec{x}^-\in\mathbb{R}^n} \quad \sum_{i=1}^n (x_i^+ + x_i^-) \tag{9.10}$$

$$\text{s.t.} \quad A(\vec{x}^+ - \vec{x}^-) = \vec{y} \tag{9.11}$$

$$\vec{x}^+ \geq \vec{0} \tag{9.12}$$

$$\vec{x}^- \geq \vec{0}. \tag{9.13}$$

This is a linear program which is efficiently solvable.

As a corollary, we can consider the ℓ^1-norm regression problem:

$$\min_{\vec{x}\in\mathbb{R}^n} \|A\vec{x} - \vec{y}\|_1 . \tag{9.14}$$

We can introduce the slack variable $\vec{e} = A\vec{x} - \vec{y}$ and obtain the problem:

$$\min_{\substack{\vec{x}\in\mathbb{R}^n \\ \vec{e}\in\mathbb{R}^m}} \quad \|\vec{e}\|_1 \tag{9.15}$$

$$\text{s.t.} \quad A\vec{x} - \vec{y} = \vec{e} \tag{9.16}$$

which is an equality-constrained ℓ^1 minimization problem, and thus a linear program as demonstrated above.

Example 210 (Mean Versus Median)**.** Let k be a positive integer. Suppose we have points $\vec{x}_1, \ldots, \vec{x}_k \in \mathbb{R}^n$. Consider the problem

$$\min_{\vec{x}\in\mathbb{R}^n} \sum_{i=1}^{k} \|\vec{x} - \vec{x}_i\|_2^2 . \tag{9.17}$$

This is an unconstrained strongly convex differentiable problem, so it has a unique solution $\vec{x}_1^\star$ which we may find by setting the derivative of the objective to $\vec{0}$. We obtain

$$\vec{0} = 2\sum_{i=1}^{k}(\vec{x}_1^\star - \vec{x}_i) \tag{9.18}$$

$$\implies \vec{0} = \sum_{i=1}^{k}(\vec{x}_1^\star - \vec{x}_i) = k \cdot \vec{x}_1^\star - \sum_{i=1}^{k}\vec{x}_i \tag{9.19}$$

$$\implies \vec{x}_1^\star = \frac{1}{k}\sum_{i=1}^{k}\vec{x}_i. \tag{9.20}$$

This computation implies that the sample mean is the point which minimizes the total squared distance to all points in the dataset.

Now suppose that we instead consider the problem

$$\min_{\vec{x}\in\mathbb{R}^n} \sum_{i=1}^{k} \|\vec{x} - \vec{x}_i\|_2 . \tag{9.21}$$

The solution to this problem is the *sample median* of the points. To see this, suppose that $n = 1$, i.e., all our data x_i are scalar-valued. Then we obtain the problem

$$\min_{x\in\mathbb{R}} \sum_{i=1}^{k} |x - x_i| . \tag{9.22}$$

This is an unconstrained, convex, non-differentiable problem. Let us examine all critical points – that is, points where the derivative is 0 or undefined. The derivative of the objective is

$$\frac{\mathrm{d}}{\mathrm{d}x}\sum_{i=1}^{k}|x - x_i| = \sum_{i=1}^{k}\frac{\mathrm{d}}{\mathrm{d}x}|x - x_i| \tag{9.23}$$

$$= \sum_{i=1}^{k}\begin{cases} 1, & \text{if } x > x_i \\ -1, & \text{if } x < x_i \\ \text{undefined}, & \text{if } x = x_i \end{cases} \tag{9.24}$$

$$= \begin{cases} \sum_{i:\ x>x_i} 1 + \sum_{i:\ x<x_i} -1, & \text{if } x \notin \{x_1, \ldots, x_k\} \\ \text{undefined}, & \text{if } x \in \{x_1, \ldots, x_k\} \end{cases} \tag{9.25}$$

$$= \begin{cases} |\{i \in \{1,\dots,k\} \colon x > x_i\}| - |\{i \in \{1,\dots,k\} \colon x < x_i\}|, & \text{if } x \notin \{x_1,\dots,x_k\} \\ \text{undefined}, & \text{if } x \in \{x_1,\dots,x_k\}. \end{cases} \tag{9.26}$$

Thus if x is such that $|\{i \in \{1,\dots,k\} \colon x > x_i\}| = |\{i \in \{1,\dots,k\} \colon x < x_i\}|$, then the derivative is 0, so this x is a candidate solution. To put this convoluted-looking condition in words, notice that the first term in hte equality is just the number of x_i which are larger than x, and the second term is the number of x_i which are smaller than x. Thus the condition says that there are the same number of points in the set which are larger than x as there are points which are smaller than x. This x would fulfill the traditional definition of "median" as the middle of the sorted list of points.

To formally solve this problem, one must also check all the values $x = x_i$ and compare the objective values. But eventually after doing all this, one recovers that the optimal solutions are all possible medians of the dataset.

Because the median is defined using the $|\cdot|$ instead of $(\cdot)^2$ function, it inherits several different properties. The most striking is its robustness; the median is much more robust than the mean. The mean is very sensitive to outliers, while the median is less sensitive (i.e. if we blow up an outlier point, the mean will change a lot, while the median will be unaffected).

9.3 Analysis of LASSO Regression

In this section we will solve the one-dimensional LASSO problem. The ideas generalize to the vector case directly, through a reduction of the vector LASSO problems to several one-dimensional LASSO problems. The details of this reduction are left as a homework exercise.

First consider the scalar ridge regression problem, with $\vec{a} \neq \vec{0}$:

$$\min_{x \in \mathbb{R}} f_{\text{RR}}(x) \qquad \text{where} \qquad f_{\text{RR}}(x) \doteq \frac{1}{2} \|\vec{a}x - \vec{y}\|_2^2 + \frac{1}{2}\lambda x^2. \tag{9.27}$$

By taking the derivative, we get

$$\frac{\mathrm{d}f_{\text{RR}}}{\mathrm{d}x}(x) = \vec{a}^\top(\vec{a}x - \vec{y}) + \lambda x \tag{9.28}$$

$$= (\vec{a}^\top\vec{a} + \lambda)x - \vec{a}^\top\vec{y} \tag{9.29}$$

$$= (\|\vec{a}\|_2^2 + \lambda)x - \vec{a}^\top\vec{y} \tag{9.30}$$

$$\implies x^\star_{\text{RR}} = \frac{\vec{a}^\top\vec{y}}{\|\vec{a}\|_2^2 + \lambda}. \tag{9.31}$$

By setting $\lambda = 0$ we obtain the least squares solution:

$$x^\star_{\text{LS}} = \frac{\vec{a}^\top\vec{y}}{\|\vec{a}\|_2^2} \tag{9.32}$$

Now consider the scalar LASSO problem

$$\min_{x \in \mathbb{R}} f_{\text{LASSO}}(x) \qquad \text{where} \qquad f_{\text{LASSO}}(x) \doteq \frac{1}{2} \|\vec{a}x - \vec{y}\|_2^2 + \lambda |x| . \tag{9.33}$$

We simplify the objective, obtaining

$$f_{\text{LASSO}}(x) = \frac{1}{2}\|\vec{a}\|_2^2 x^2 - (\vec{a}^\top\vec{y})x + \frac{1}{2}\|\vec{y}\|_2^2. \tag{9.34}$$

We first claim that if $x^\star_{\text{LASSO}} \neq 0$, then $\text{sign}(x^\star_{\text{LASSO}}) = \text{sign}(\vec{a}^\top\vec{y})$. This is true because if $\text{sign}(x^\star_{\text{LASSO}}) = \text{sign}(\vec{a}^\top\vec{y})$ then the second term $(\vec{a}^\top\vec{y})x^\star_{\text{LASSO}}$ is negative (thus making the objective as small as possible), whereas if

$\operatorname{sign}(x^\star_{\text{LASSO}}) = -\operatorname{sign}(\vec{a}^\top \vec{y})$ then the second term $(\vec{a}^\top \vec{y})x^\star_{\text{LASSO}}$ is positive (thus making the objective *not* as small as possible). We will use this fact slightly later.

Now f_{LASSO} has a derivative everywhere except $x = 0$. We obtain

$$\frac{\mathrm{d}f_{\text{LASSO}}}{\mathrm{d}x}(x) = \vec{a}^\top(\vec{a}x - \vec{y}) + \lambda \begin{cases} 1, & \text{if } x > 0 \\ -1, & \text{if } x < 0 \\ \text{undefined}, & \text{if } x = 0. \end{cases} \tag{9.35}$$

Let $x^\star$ be a critical point of this problem. We solve what $x^\star$ should be using casework.

Case 1. If $x^\star_{\text{LASSO}} > 0$, then the derivative is well-defined, so it must be equal to 0. Thus we have

$$\begin{aligned} 0 &= \frac{\mathrm{d}f_{\text{LASSO}}}{\mathrm{d}x}(x^\star_{\text{LASSO}}) && (9.36)\\ &= \vec{a}^\top(\vec{a}x^\star_{\text{LASSO}} - \vec{y}) + \lambda && (9.37)\\ \implies x^\star_{\text{LASSO}} &= \frac{\vec{a}^\top\vec{y} - \lambda}{\|\vec{a}\|_2^2}. && (9.38) \end{aligned}$$

Thus if $x^\star_{\text{LASSO}} > 0$ then $x^\star_{\text{LASSO}} = \frac{\vec{a}^\top\vec{y}-\lambda}{\|\vec{a}\|_2^2}$. As a corollary, if $x^\star_{\text{LASSO}} > 0$ then $\vec{a}^\top\vec{y} > \lambda$.

Case 2. If $x^\star_{\text{LASSO}} < 0$, then the derivative is well-defined, so it must be equal to 0. Thus we have

$$\begin{aligned} 0 &= \frac{\mathrm{d}f_{\text{LASSO}}}{\mathrm{d}x}(x^\star_{\text{LASSO}}) && (9.39)\\ &= \vec{a}^\top(\vec{a}x^\star_{\text{LASSO}} - \vec{y}) - \lambda && (9.40)\\ \implies x^\star_{\text{LASSO}} &= \frac{\vec{a}^\top\vec{y} + \lambda}{\|\vec{a}\|_2^2}. && (9.41) \end{aligned}$$

Thus if $x^\star_{\text{LASSO}} < 0$ then $x^\star_{\text{LASSO}} = \frac{\vec{a}^\top\vec{y}+\lambda}{\|\vec{a}\|_2^2}$. As a corollary, if $x^\star_{\text{LASSO}} < 0$ then $\vec{a}^\top\vec{y} < -\lambda$.

Case 3. $x^\star = 0$. The above two cases have shown that if $x^\star_{\text{LASSO}} \neq 0$ then $|\vec{a}^\top\vec{y}| > \lambda$. The contrapositive of this, which must also be true, is that if $|\vec{a}^\top\vec{y}| \leq \lambda$ then $x^\star_{\text{LASSO}} = 0$.

Since we have computed $x^\star_{\text{LASSO}}$ in all cases and know it is nonzero except in the third case, we can strengthen the "if" statement in the third case to an if-and-only-if — namely, $|\vec{a}^\top\vec{y}| \leq \lambda$ if and only if $x^\star_{\text{LASSO}} = 0$. Using this in conjunction with the above fact that if $x^\star_{\text{LASSO}} \neq 0$ then $\operatorname{sign}(x^\star_{\text{LASSO}}) = \operatorname{sign}(\vec{a}^\top\vec{y})$, we obtain that $x^\star_{\text{LASSO}} > 0$ if and only if $\vec{a}^\top\vec{y} > \lambda$ and $x^\star_{\text{LASSO}} < 0$ if and only if $\vec{a}^\top\vec{y} < -\lambda$. Using our findings from the first two cases, we obtain the following trichotomy:

- $x^\star_{\text{LASSO}} > 0 \iff \vec{a}^\top\vec{y} > \lambda$, in which case $x^\star = (\vec{a}^\top\vec{y} - \lambda)/\|\vec{a}\|_2^2$;
- $x^\star_{\text{LASSO}} < 0 \iff \vec{a}^\top\vec{y} < -\lambda$, in which case $x^\star = (\vec{a}^\top\vec{y} + \lambda)/\|\vec{a}\|_2^2$; and
- $x^\star_{\text{LASSO}} = 0 \iff -\lambda \leq \vec{a}^\top\vec{y} \leq \lambda$,

or in other words,

$$x^\star_{\text{LASSO}} = \begin{cases} (\vec{a}^\top\vec{y} - \lambda)/\|\vec{a}\|_2^2, & \text{if } \vec{a}^\top\vec{y} > \lambda \\ (\vec{a}^\top\vec{y} + \lambda)/\|\vec{a}\|_2^2, & \text{if } \vec{a}^\top\vec{y} < -\lambda \\ 0, & \text{if } -\lambda \leq \vec{a}^\top\vec{y} \leq \lambda. \end{cases} \tag{9.42}$$

As a function of $\vec{a}^\top \vec{y}$, the solution $x^\star_{\text{LASSO}}$ looks like:

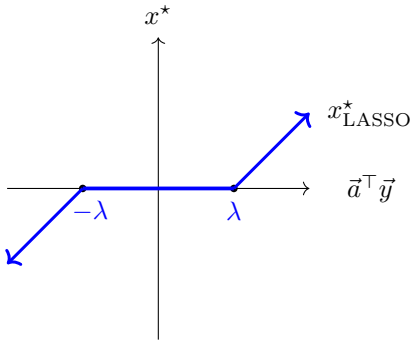

Figure 9.2: The plot of the function which maps $\vec{a}^\top \vec{y} \mapsto x^\star_{\text{LASSO}}$, where the latter term is the solution to our scalar LASSO problem. When $\vec{a}^\top \vec{y} \in [-\lambda, \lambda]$, we have $x^\star = 0$. The function $\vec{a}^\top \vec{y} \mapsto \vec{x}^\star$ is continuous, yet not differentiable at $\vec{a}^\top \vec{y} = \pm\lambda$.

If we plot the least squares solution in red on the same graph, it has the same nonzero slope as the LASSO solution, and looks like this:

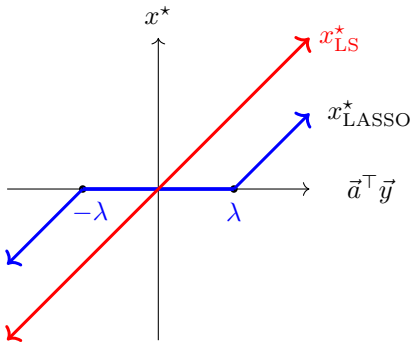

Figure 9.3: In red, we add the plot of the function which maps $\vec{a}^\top \vec{y} \mapsto x^\star_{\text{LS}}$, where the latter term is the solution to to our scalar least squares problem. Note that the LASSO solution (in blue) is always closer to zero than the least squares solution, and is set directly to zero when $\vec{a}^\top \vec{y} \in [-\lambda, \lambda]$.

This illustrates a concept called *soft thresholding*: in the regime where the least squares solution $x^\star_{\text{LS}}$ is already close to zero, $x^\star_{\text{LASSO}}$ becomes exactly zero. Meanwhile, ridge regression does not do this: $x^\star_{\text{RR}} = 0$ if and only if $\vec{a}^\top \vec{y} = 0$, which is exactly when the unregularized least squares solution itself is zero. This fundamental difference is why the solutions to LASSO regression tend to be sparse, i.e., have many entries set to 0.

9.4 Geometry of LASSO Regression

In this section we introduce a geometric description of LASSO. The geometry relies on a crucial theorem which unveils a deep connection between regularization and constrained optimization for convex problems. We state the result here; the proof uses duality theory and is left to homework.

Theorem 211

Let $f_0 \colon \mathbb{R}^n \to \mathbb{R}$ be strictly convex and such that $\lim_{t\to\infty} f_0(\vec{x}_t) = \infty$ for all sequences $(\vec{x}_t)_{t=0}^\infty$ such that $\lim_{t\to\infty} \|\vec{x}_t\|_2 = \infty$,[a] and $R \colon \mathbb{R}^n \to \mathbb{R}_+$ be convex and take non-negative values. Further suppose that there

exists $\vec{x}_0 \in \mathbb{R}^n$ such that $R(\vec{x}_0) = 0$.

For $\lambda \geq 0$ and $k \geq 0$, let $\mathcal{R}(\lambda)$ and $\mathcal{C}(k)$ be sets of solutions to the "regularized" and "constraint" programs:

$$\mathcal{R}(\lambda) \doteq \operatorname*{argmin}_{\vec{x} \in \mathbb{R}^n} \{f_0(\vec{x}) + \lambda R(\vec{x})\} \tag{9.43}$$

$$\mathcal{C}(k) \doteq \operatorname*{argmin}_{\substack{\vec{x} \in \mathbb{R}^n \\ R(\vec{x}) \leq k}} f_0(\vec{x}). \tag{9.44}$$

Then:

(a) for every $\lambda \geq 0$ there exists $k \geq 0$ such that $\mathcal{R}(\lambda) = \mathcal{C}(k)$; and

(b) for every $k > 0$ there exists $\lambda \geq 0$ such that $\mathcal{R}(\lambda) = \mathcal{C}(k)$.

[4] This assumption is called "coercivity".

This shows that in some sense, regularized convex problems are equivalent to constrained convex problems; and in this equivalence, the regularizer for the regularized problem shapes the constraint set of the constrained problem. In particular, regularized least squares ($f_0(\vec{x}) = \|A\vec{x} - \vec{y}\|_2^2$) with full column rank is equivalent to constrained least squares (with the same f_0).

Now, we sketch the feasible sets and level sets of the objective function for the constrained problems corresponding to both ridge regression and LASSO regression.

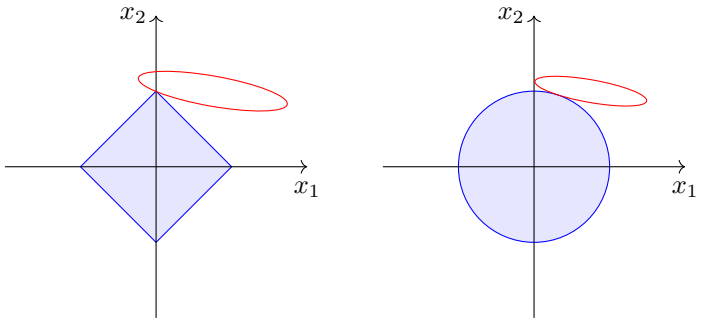

Figure 9.4: Geometric differences between LASSO and ridge regression. On the left side, the blue diamond depicts the feasible region for an ℓ^1-norm constraint such as $\|\vec{x}\|_1 \leq t$, while the circle on the right side is the feasible region for an ℓ^2-norm constraint such as $\|\vec{x}\|_2^2 \leq t$. On both graphs, the red line is a level set of our objective function; specifically, the minimal level set that still intersects the feasible region. The intersection of this level set with the feasible region is the solution to our constrained problem and thus to an equivalent regularized problem.

Note how with the ℓ^1-norm constraint, the intersection of the feasible region with the minimal level set is more likely to be at a corner of the feasible region, which is a point where some coordinates are set exactly to zero. Meanwhile, with the ℓ^2-norm constraint, the intersection can be at an arbitrary point on the circle (or sphere in higher dimensions), and likely isn't at a corner. This is why LASSO induces sparsity in $\vec{x}$, due to the distinctive corners we saw earlier in its norm ball. Meanwhile, although ridge regression compresses $\vec{x}^\star_{\text{RR}}$ to be smaller, it doesn't necessarily induce sparsity in $\vec{x}^\star_{\text{RR}}$.