Lecture 21. Strong duality. Slater's condition.

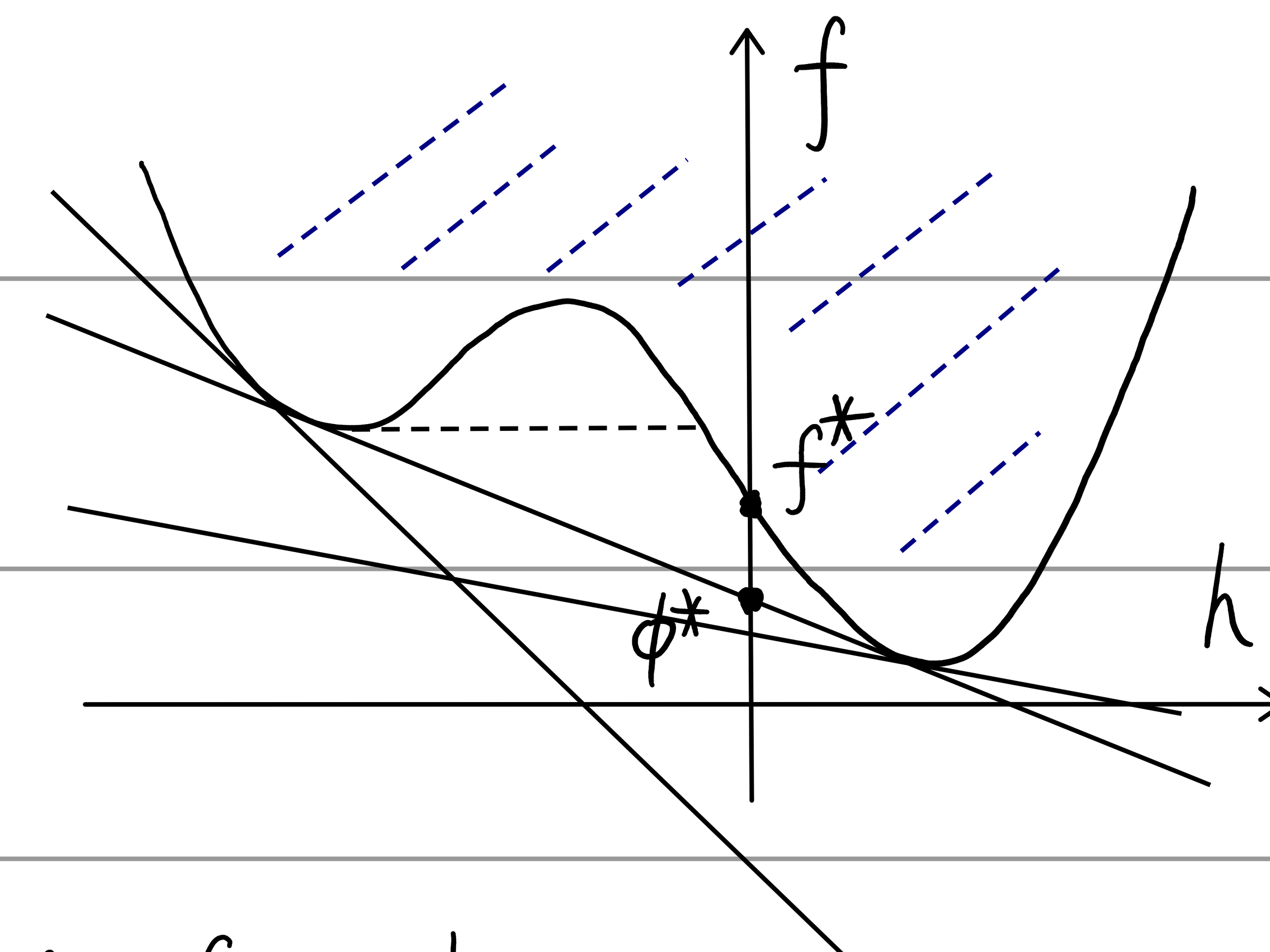
Lagrange dual: $\phi(\lambda, \mu) = \inf_{x \in D} L(x, \lambda, \mu)$. $\sup_{\mu \geq 0} \phi(\lambda, \mu)$.

Strong duality: $f^* = \inf_{x \in \Omega} f(x) = \sup_{\mu \geq 0} \phi(\lambda, \mu) = \phi^*$.

Slater's condition for convex optimization:

If $x \in \text{relint}(D)$ such that $\begin{cases} g_i(x) = 0 \\ h_j(x) < 0 \end{cases}$ then strong duality holds.

Geometric interpretation of duality.



min $x^4 - 50x^2 + 100x$.
s.t. $x \geq -5/2$. $(h(x) = -5/2 - x)$.

consider the epigraph of $(h(x), f(x))$.

$C = \{(p, t) : \exists x.\ h(x) \leq p.\ f(x) \leq t\}$. $l: f + \mu h = y$. $\mu \geq 0$.

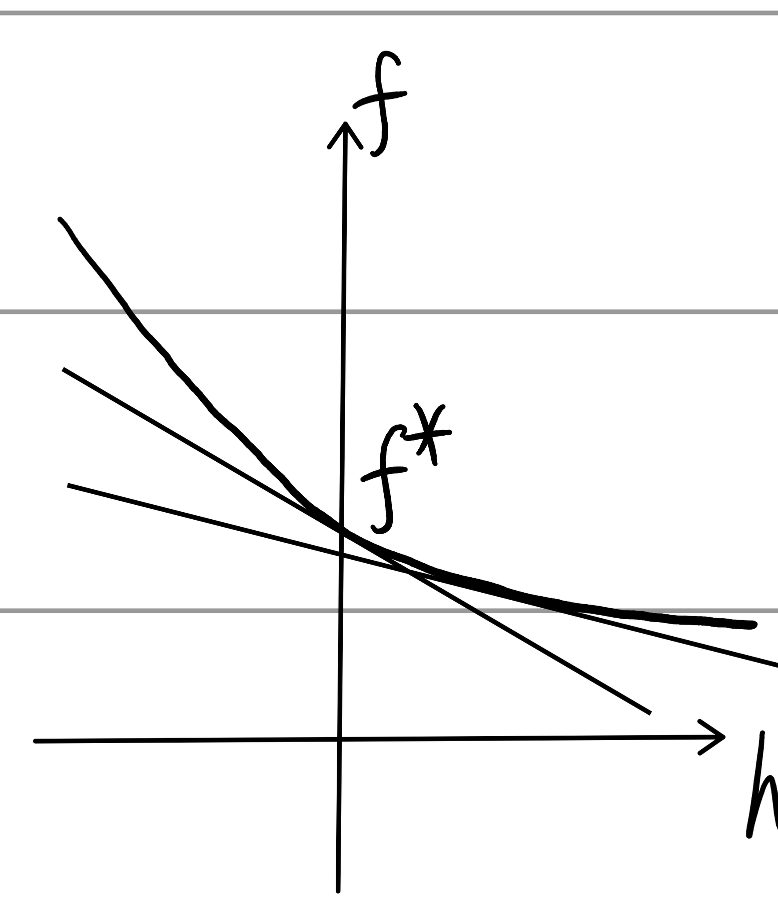
$\phi(\mu) = \inf_x f(x) + \mu \cdot h(x) = \inf_x y : \exists x.\ f(x) + \mu h(x) = y$.

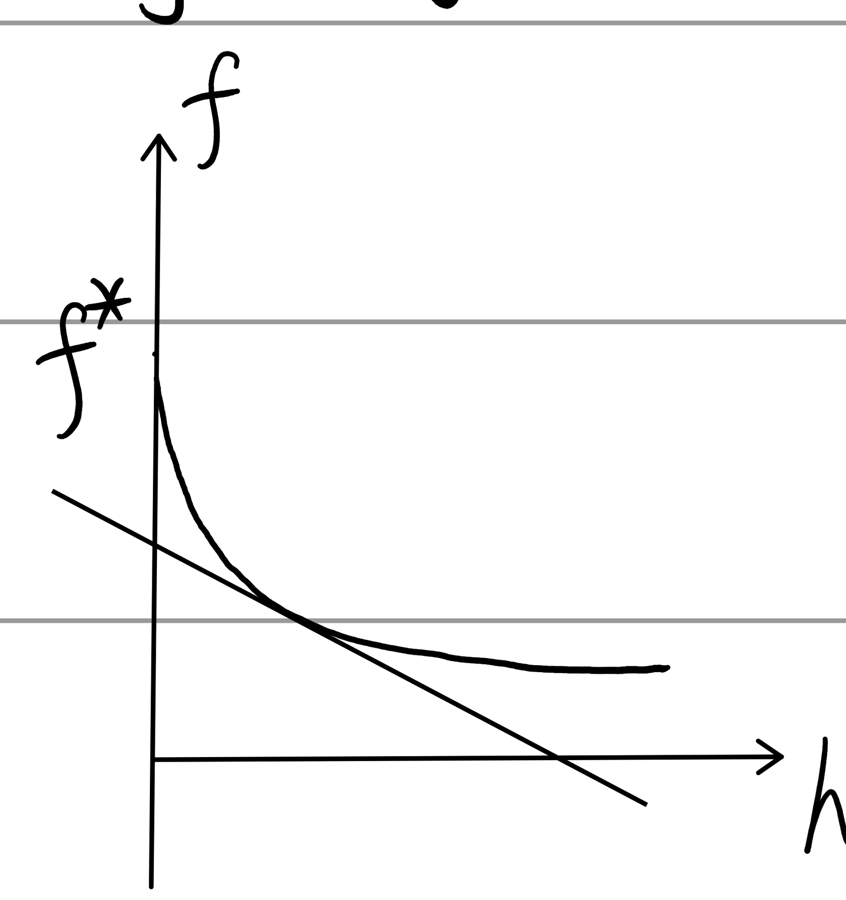
$\phi(\mu)$ = the smallest y of lines : $f + \mu h = y$ that intersects graph (h, f)

l is a supporting hyperplane of C. f^*: C intersects axis
ϕ^*: l intersects axis

Strong duality does not hold since there is no SH passing through $(0, f^*)$

However, if C is convex. for any points $\in \partial C$. there is a SH passing through it.



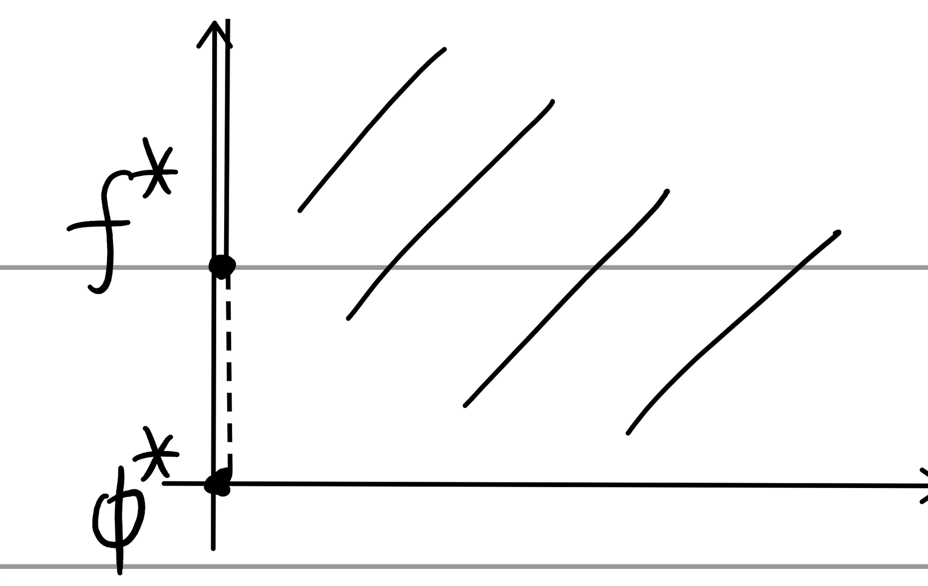


Is it good enough? No. strong duality doesn't always hold for convex.

when does this argument fail? The unique supporting hyperplane is vertical.

Recall the example $\min_{x,y>0} e^{-x}$ s.t. $x^2/y \le 0$.

$C = \{(0,t) : t \ge 1\} \cup \{(p,t) : p>0, t>0\}$



Intuitively, if we have Slater's condition, then $\exists$ nonvertical SH.

Definition: For a convex problem $\min f(x)$. s.t. $g_1(x) = \cdots = g_m(x) = 0$, $h_1(x) \le 0, \cdots h_l(x) \le 0$

define $C = \{(p,q,t) : \exists x \in D.\ h_j(x) \le p_j,\ g_i(x) = q_i,\ f(x) \le t\}$.

Lemma. C is convex.

Proof. Take two points $(p_1, q_1, t_1), (p_2, q_2, t_2) \in C$. So

$\left.\begin{array}{l} \exists x_1 \text{ s.t. } h(x_1) \le p_1.\ g(x_1) = q_1.\ f(x_1) \le t. \\ \exists x_2 \text{ s.t. } h(x_2) \le p_2.\ g(x_2) = q_2.\ f(x_2) \le t. \end{array}\right\}$ By convexity of f and h. affinity of g.

$\forall \theta \in [0,1]$. let $y = \theta x_1 + \bar{\theta} x_2$. then we have

$$\left.\begin{array}{l} h(y) \le \theta h(x_1) + \bar{\theta} h(x_2) \le \theta p_1 + \bar{\theta} p_2 \\ g(y) = \theta g(x_1) + \bar{\theta} g(x_2) = \theta q_1 + \bar{\theta} q_2 \\ f(y) \le \theta f(x_1) + \bar{\theta} f(x_2) \le \theta t_1 + \bar{\theta} t_2 \end{array}\right\} \Rightarrow \theta(p_1, q_1, t_1) + \bar{\theta}(p_2, q_2, t_2) \in C.$$

$\square$.

Proof of strong duality under Slater's condition.

If $f^* = -\infty$. by weak duality. $\phi^* \le f^*$. So $\phi^* = -\infty$. strong holds.

Now assume $f^* > -\infty$. By slater's condition. feasible set $\Omega \neq \phi$. So $f^* < \infty$. It suffices to show $\exists$ a nonvertical SH passing through $(0,0,f^*)$

Step 1. $(0,0,f^*) \in \partial C$. (note that $(0,0,f^*)$ may not in C. it is in $\mathrm{cl}\, C$)

$f^* = \inf_{x \in \Omega} f(x)$. So $\forall \varepsilon > 0$. $\exists x$. $g(x) = 0$. $h(x) \leq 0$. $f(x) < f^* + \varepsilon$. Thus $(0,0,f^*+\varepsilon) \in C$. which implies $(0,0,f^*) \in \mathrm{cl}\, C$. Moreover. $\forall \delta > 0$. $\nexists x \in \Omega$. $f(x) \leq f^* - \delta$. So $(0,0,f^*-\delta) \notin C$. Thus $(0,0,f^*) \notin \mathrm{int}\, C$.

Step 2. $(0,0,f^*) \in \partial C$. $\Rightarrow$ $\exists$ supporting hyperplane passing through it.

i.e. $\exists (\mu, \lambda, \xi) \neq \vec{0}$. s.t. $\forall (p, q, t) \in C$, $\mu^T p + \lambda^T q + \xi t \geq \xi f^*$.

Since $\forall t > f^*$. $(0,0,t) \in C$. we have $\xi \geq 0$.

$\forall p > \vec{0}$. $(p, q, f^*) \in C$. we have $\mu \geq \vec{0}$.

We now show that $\xi \neq 0$. otherwise. $\mu^T p + \lambda^T q \geq 0$.

By Slater's condition. $\exists \tilde{x}$. s.t. $g(\tilde{x}) = \vec{0}$. $h(\tilde{x}) < \vec{0}$. So $\exists \tilde{t}$ s.t. $(h(\tilde{x}), g(\tilde{x}), \tilde{t}) \in C$. Thus $\mu^T h(\tilde{x}) \geq 0$. Since $h(\tilde{x}) < \vec{0}$. we have $\mu = \vec{0}$. Now we obtain $\lambda^T g(x) \geq 0$ for all $x \in D$.

Also $\lambda \neq \vec{0}$ as $(\mu, \lambda, \xi) \neq \vec{0}$. Let $g(x) = Ax - b$. where A has full rank. By Slater's condition. $\tilde{x} \in \mathrm{int}\, D$. $g(\tilde{x}) = \vec{0}$. Thus $\exists \hat{x} \in D$

$\lambda^T g(\hat{x}) < 0$. contradiction. Finally we conclude $\xi > 0$.

Step 3. Now let $\tilde{\mu} = \mu/\xi$, $\tilde{\lambda} = \lambda/\xi$. Then $\forall (p, q, t) \in C$, we have $\tilde{\mu}^T p + \tilde{\lambda}^T q + t \geq f^*$. $\phi(\tilde{\lambda}, \tilde{\mu}) = \inf_{x \in D} f(x) + \tilde{\lambda}^T g(x) + \tilde{\mu}^T h(x)$ $\geq f^*$ (since $(h(x), g(x), f(x)) \in C$ if $x \in D$). $\Rightarrow \phi(\tilde{\lambda}, \tilde{\mu}) \geq f^*$

Note that $\tilde{\mu} \geq \vec{0}$. So $\phi^* = \sup_{\mu \geq 0} \phi(\lambda, \mu) \geq \phi(\tilde{\lambda}, \tilde{\mu})$. $\Rightarrow \phi^* = f^*$. $\square$

Remark. Can relax $h_j(\tilde{x}) < 0$ by $h_j(\tilde{x}) \leq 0$ if h_j is affine.

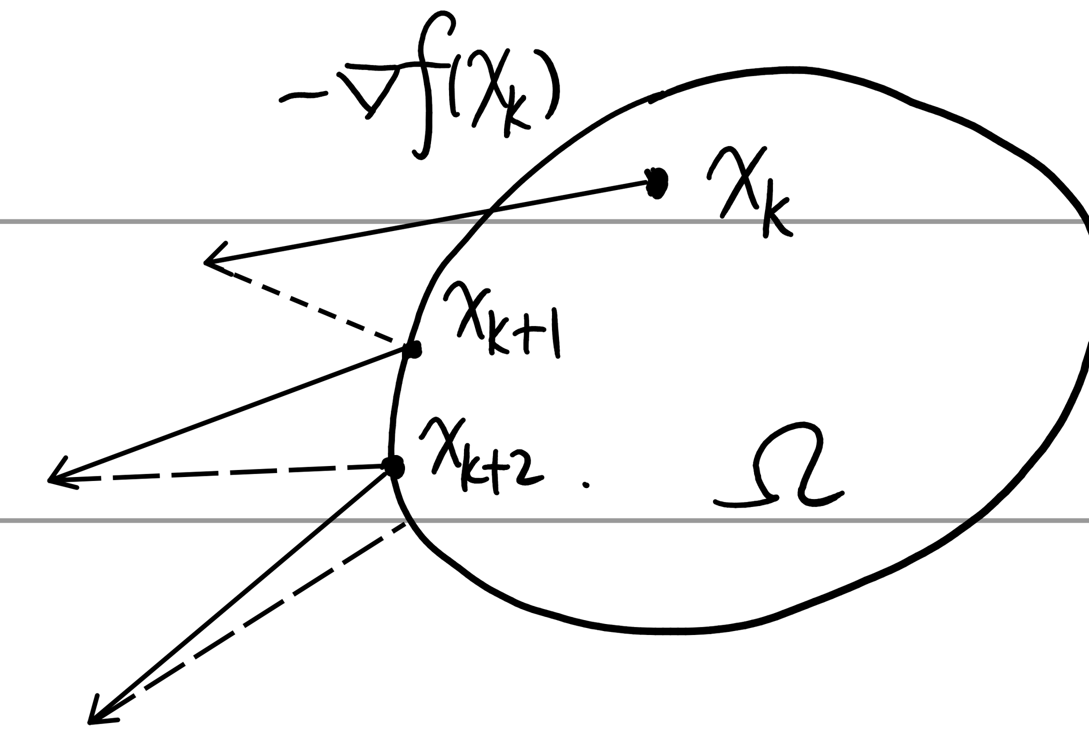
Solving inequality constrained problems: projected gradient descent

Recall the gradient descent method: $x_{k+1} = x_k + t_k d_k$. $= -\nabla f(x_k)$

Now we consider $\min_{x \in \Omega} f(x)$.

If $x_k - t\nabla f(x_k)$ is infeasible.

project it onto Ω.



Projection operator: $P_\Omega(y) = \arg\min_{x \in \Omega} \|x - y\|$.

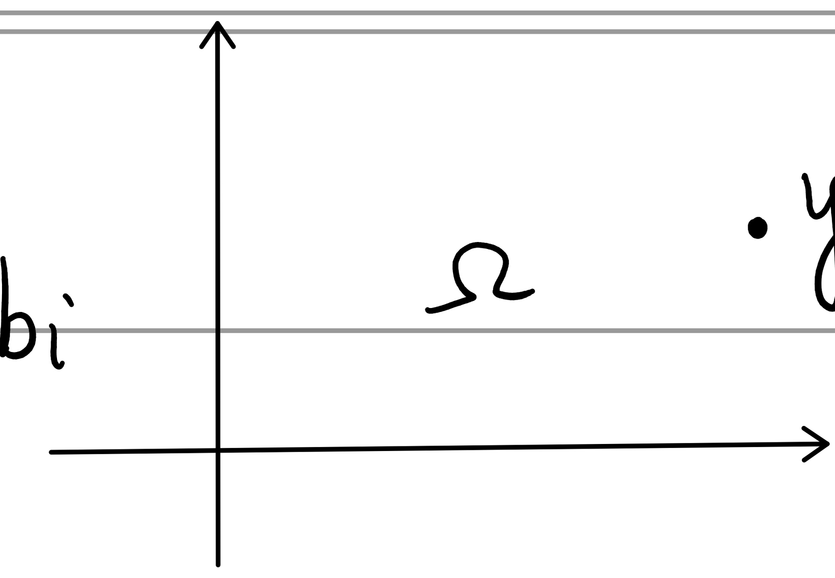
Projected gradient descent: $x_{k+1} = x_k - P_\Omega(x_k - t\nabla f(x_k))$.

In general, projection is an optimization which is hard to solve.

Here we discuss some examples where projection can be efficiently computed

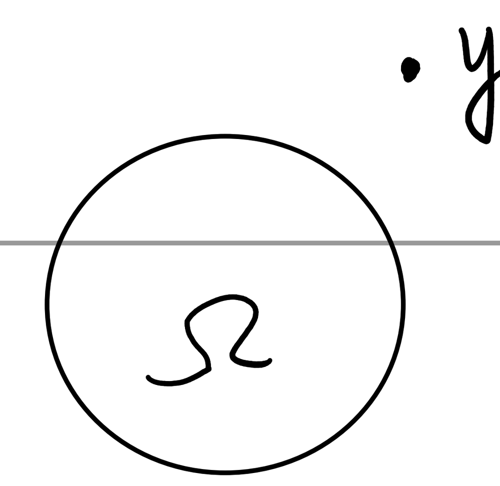
Example. Box constraints. $\Omega = \{x : a_i \leq x_i \leq b_i,\ i = 1, 2, \dots, n\}$.

$$(P_\Omega(y))_i = \min\{b_i, \max\{a_i, y_i\}\} = \begin{cases} a_i & y_i < a_i \\ y_i & a_i \le y_i \le b_i \\ b_i & y_i > b_i \end{cases}$$



Example. l_2 constraints. $\Omega = \{x : \|x\|_2 \le t\}$ (ridge regression).

$P_\Omega(y) = \min\{1, \frac{t}{\|y\|_2}\} y$. why?



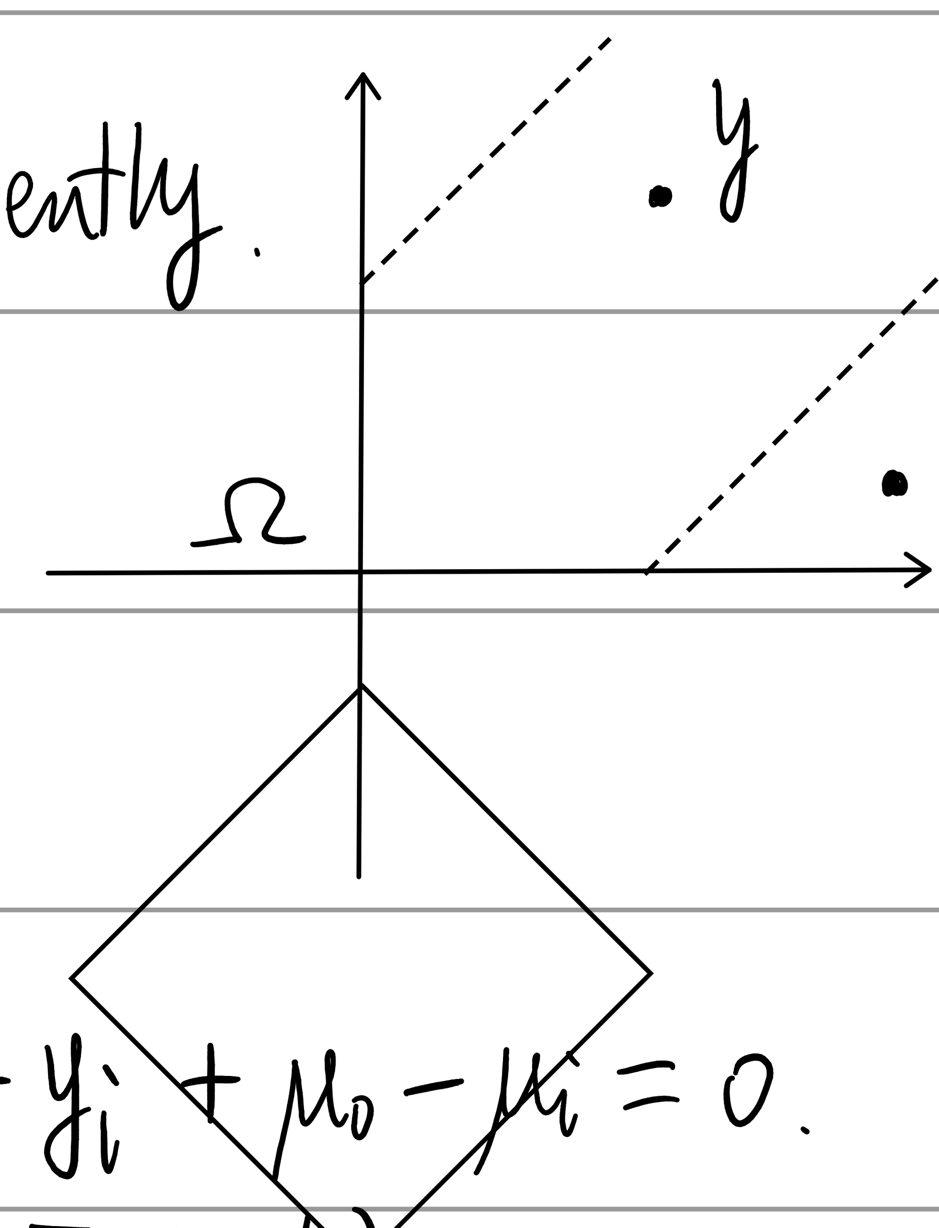
$\min \|x-y\|^2$. s.t. $\|x\|_2^2 \le t^2$. By KKT condition.

$\exists \mu \ge 0$. s.t. $2(x-y) + 2\mu x = 0$. $\Rightarrow y = (1+\mu)x \propto x$.

$\mu(\|x\|^2 - t) = 0$. either $\mu = 0$ or $\|x\|_2 = t$.

Example: l_1 constraints. $\Omega = \{x : \|x\|_1 \le t\}$ (lasso).

no closed form. but can be computed efficiently.



By symmetry. only need to consider $y \ge 0$.

$\min \|x-y\|^2$. s.t. $\sum x_i \le t$. $x_i \ge 0$ $\forall i$.

By KKT conditions. $\exists \mu_0, \mu_1, \dots, \mu_n \ge 0$. $\begin{cases} x_i - y_i + \mu_0 - \mu_i = 0. \\ \mu_0(\sum x_i - t) = 0 \\ \mu_i x_i = 0. \end{cases}$

Case 1. $\|y\|_1 \le t \Rightarrow \begin{matrix} \mu_0 = 0 \\ \mu_i = 0. \end{matrix}$ $x = y$.

Case 2. $\|y\|_1 > t. \Rightarrow \mu_0 > 0$. $x_i = y_i - \mu_0 + \mu_i$. $\mu_i x_i = 0$.

if $\mu_i = 0 \Rightarrow x_i = y_i - \mu_0 \ge 0$. if $x_i = 0 \Rightarrow \mu_i = \mu_0 - y_i \ge 0$.

$\Rightarrow x_i = \begin{cases} y_i - \mu_0 & \text{if } y_i \ge \mu_0 \\ 0 & \text{o.w.} \end{cases}$ s.t. $\sum x_i = t$.

w.l.o.g. assume $y_1 \ge y_2 \ge \dots \ge y_n$. binary search.