Lecture 4 Dual programs

- conjugate function
- examples of conjugate functions
- convex programs and examples
- dual programs and weak duality
- examples

Conjugate functions

Goal: defining the dual function of a function $f: \mathbb{R}^n \to \mathbb{R}$

the dual function has to be of the form $f^*: \mathbb{R}^n \to \mathbb{R}$ and (as for sets) has to maintain the information about f is f is convex. (for simplicity we assume that $f(x) > -\infty \ \forall x$)

What to do?

supporting hyperplanes of epi f

$$\text{epi} f = \{(x,t) : x \in \text{dom} f, \ t \geq f(x)\}$$

$$S_{\text{epi} f}(y, \mu) = \sup \{ y^T x + \mu t \mid x \in \text{dom} f, \ t \geq f(x) \}$$

if $\mu > 0$ then $S_{\text{epi} f}(y,\mu) = +\infty$ (t is unbounded)

So to get useful information, we only need to look at $\mu \leq 0$

$\Longrightarrow$

$\mu \leq 0$

$$S_{epif}(y,\mu) = \sup\left\{ y^T x + \mu t \,\middle|\, \begin{array}{l} x \in \mathrm{dom} f \\ t \geq f(x) \end{array} \right\} \underset{=}{\mu \leq 0}$$

$$= \sup\{ y^T x + \mu f(x) \mid x \in \mathrm{dom} f \}$$

$\mu = 0 \rightsquigarrow$ does not store any usefull information about the function value

$\mu < 0$ then we can rescale (y,μ) since the usefull information is stored in $\left(\frac{y}{-\mu}, -1\right)$

Overall: we just care about tuples of the form $(y,-1)$

Definition of conjugate function:

Given $f: \mathbb{R}^n \to \mathbb{R}$ the conjugate function f^* is defined as

$$f^*(y) = \sup\{ y^T x - f(x) \mid x \in \mathrm{dom} f \}$$

Observation: f^* is convex (pointwise supremum of affine (in y) functions)

Theorem

f is convex
epi f is a closed convex set $\Big\} \Rightarrow f^{**} = f$

examples of conjugate functions

$f(x) = x \log x$, $x \in \mathbb{R}^+$ (negative entropy)

$$f^*(y) = \sup_{x>0} (y \cdot x - x \log x)$$

$$(yx - x\log x)' = 0 \Rightarrow y - 1 - \log x \Rightarrow$$
$$\Rightarrow x = e^{y-1}$$

$$f^*(y) = y \cdot e^{y-1} - e^{y-1} \log e^{y-1} =$$
$$= y \cdot e^{y-1} - e^{y-1}(y-1) = e^{y-1}$$

$f(x) = \frac{1}{2} x^T Q x$, Q invertible and symmetric $f^*(y) = \sup_x (y^T x - \frac{1}{2} x^T Q x)$

$$\left(y^T x - \frac{1}{2} x^T Q x\right)' = 0 \Rightarrow y - Qx = 0 \Rightarrow$$

$$\Rightarrow x = Q^{-1} y$$

$$\Rightarrow f^*(y) = y^T Q^{-1} y - \frac{1}{2} y^T (Q^{-1})^T Q (Q^{-1} y) =$$
$$= y^T Q^{-1} y - \frac{1}{2} y^T Q^{-1} y = \frac{1}{2} y^T Q^{-1} y$$

Convex programs

$$\inf \; f_0(x)$$
$$\text{s.t} \quad f_i(x) \leq 0 \quad \forall i \in [m]$$
$$h_j(x) = 0 \quad \forall j \in [p]$$

convex functions

$\hookrightarrow$ affine functions $(h_i(x) = \langle a_i, x \rangle - b_i)$

domain of the optimization problem D

$$D = \operatorname{dom} f_0 \cap \bigcap_{i=1}^{m} \operatorname{dom} f_i \cap \bigcap_{j=1}^{p} \operatorname{dom} h_j$$

the optimal value p^* is the value of $\inf f_0(x)$ over x that satisfies the constraints

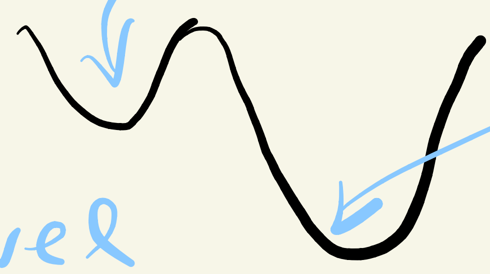
If the problem is infeasible then we define $p^* = +\infty$. If it is unbounded below $p^* = -\infty$.

(arguably) the most famous property of a convex program: <u>any local optimal solution is a global optimal solution</u>

proof

Intuition:

valley of local optimum



deeper valley of global optimum

In order to travel from the shallow valley to the deep valley I need to climp a hill and then descend $\Rightarrow$ non convex

let x be a local optimal solution $\Rightarrow$

$$\Rightarrow f_0(x) = \inf\{f_0(z) \mid z \text{ feasible}, \|z-x\|_2 \le R\}$$

for some $R > 0$

towards contradiction suppose that $\exists y \in \mathbb{R}^n$ s.t. $f_0(y) < f_0(x)$

$$\text{then } f_0(\underbrace{(1-\varepsilon)x + \varepsilon y}_{\substack{z \text{ is feasible} \\ \text{(on the line between two} \\ \text{feasible points)}}}) \overset{\text{convexity}}{\le} (1-\varepsilon)f(x) + \varepsilon f(y) < (1-\varepsilon)f(x) + \varepsilon f(x) = f(x)$$

ε small enough $\|z-x\|_2 \le R$ and $f_0(z) < f_0(x)$ which is a contradiction $\blacksquare$

differentiable functions

easy to check condition for the optimality of a solution.

x is optimal iff $\langle \nabla f_0(x), y-x \rangle \geq 0$ $\forall$ feasible solution y (Intuition: if I go versus any feasible direction, I increase)

proof

$\Leftarrow$

$f_0(y) \geq f_0(x) + \langle \nabla f_0(x), y-x \rangle$

$\langle \nabla f_0(x), y-x \rangle \geq 0$ $\forall y$ feasible $\Longrightarrow$

$\Longrightarrow$ $f_0(y) - f_0(x) \geq 0$

$\Rightarrow$

x is optimal and $\exists y$ feasible s.t

$\langle \nabla f_0(x), y-x \rangle < 0$

let $z(\varepsilon) = (1-\varepsilon)x + \varepsilon y$ ($z(\varepsilon)$ is feasible)

$g(\varepsilon) = f_0(z(\varepsilon))$, $g(0) = f_0(x)$

$g'(0) = \langle \nabla f_0(x), y-x \rangle < 0$ $\quad$ $g(\varepsilon) = g(0) + g'(0)\cdot\varepsilon + r(\varepsilon)$ $\quad$ $\lim_{\varepsilon \to 0} \frac{r(\varepsilon)}{\varepsilon} \to 0$

$\Rightarrow g(\varepsilon) < g(0)$ for small ε

corollary

for unconstrained optimization

$< \nabla f_0(x), y-x > \geq 0 \quad \forall y \in R^n \Rightarrow$

$\Rightarrow \|\nabla f_0(x)\|_2 = 0$

proof

let $y-x = t \cdot \nabla f_0(x) \quad , t < 0$

$\Rightarrow < \nabla f_0(x), y-x > = t \cdot \|\nabla f_0(x)\|^2$

$= t \|\nabla f_0(x)\|^2 \geq 0$

but since $t<0 \quad t\|\nabla f_0(x)\|^2 \leq 0 \Rightarrow$

$\Rightarrow \|\nabla f_0(x)\|^2 = 0$

Dual programs

① a way not to be constrained

② a way to prove lower bounds

we define the Lagrangian associated with $\inf f_0(x)$ as:

$$\begin{aligned} \inf \ & f_0(x) \\ \text{s.t} \ & f_i(x) \leq 0 \quad \forall i \in [m] \\ & h_i(x) = 0 \quad \forall j \in [p] \end{aligned}$$

$$\mathcal{L}(x,\lambda,\mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x)$$

Lagrangian multipliers

the vectors λ, μ dual variables

Lagrangian dual function

$$g(\lambda,\mu) = \inf_{x} \mathcal{L}(x,\lambda,\mu)$$

→ unconstrained

Observation

suppose $\lambda \geq 0$. Then $g(\lambda,\mu)$ is a lower bound on the optimal value p^* of the primal program.

proof

$\forall \tilde{x}$ feasible $\left.\begin{array}{l}\lambda_i f_i(\tilde{x}) \leq 0 \\ \mu_i f_i(\tilde{x}) = 0\end{array}\right\} \mathcal{L}(\tilde{x},\lambda,\mu) \leq f_0(\tilde{x})$

$$\Rightarrow g(\lambda,\mu) = \inf_x \mathcal{L}(x,\lambda,\mu) \leq \mathcal{L}(\tilde{x},\lambda,\mu) \leq$$
$$\leq f_0(\tilde{x})$$

and $g(\lambda,\mu) \leq f_0(\tilde{x})$ $\forall \tilde{x}$ feasible holds especially for the optimum value

$$\Rightarrow g(\lambda,\mu) \leq \inf_{x \text{ feasible}} f_0(x) \quad \forall \lambda_i \geq 0, \ \mu_i \in \mathbb{R}$$

dual program

$$\max_{\lambda \geq 0} g(\lambda,\mu) \quad \Big\} \ d^* \text{ optimal value}$$

weak duality $d^* \leq p^*$

Observation

the dual program is always a convex program no matter the primal !!!

why

$\forall x$ $\mathcal{L}(x,\lambda,\mu)$ is an affine function of λ,μ => concave of λ,μ

$g(\lambda,\mu)$ = infimum ove concave functions

=> concave

dual program => maximization of a concave function

=> convex program

examples

① least square $\min x^T x$

$$\text{s.t } Ax = b$$

$$g(\mu) = \inf_x L(x,\mu) = \inf_x \left(x^T x + \mu^T (Ax-b) \right) =$$

$$= \inf_x \left(x^T x + \mu^T A x \right) - \mu^T b$$

$$\nabla_x \left(x^T x + \mu^T A x \right) = 0 \Rightarrow$$

$$\Rightarrow 2x + A^T \mu = 0 \Rightarrow x = -\frac{1}{2} A^T \mu$$

$$g(\mu) = -\frac{1}{2} \mu^T A \left(-\frac{1}{2}\right) A^T \mu + \mu^T A \left(-\frac{1}{2}\right) A^T \mu - \mu^T b =$$

$$= \frac{1}{4} \mu^T A A^T \mu - \frac{1}{2} \mu^T A A^T \mu - \mu^T b =$$

$$= -\frac{1}{4} \mu^T A A^T \mu - \mu^T b$$

the dual is unconstrained !!!

$$\Longrightarrow$$

$$\nabla_\mu\left(-\frac{1}{4}\mu^T A A^T \mu - \mu^T b\right) = 0 \Rightarrow$$

$$\Rightarrow -\frac{1}{2} A A^T \mu - b = 0 \Rightarrow$$

$$\Rightarrow \mu = -2 (AA^T)^{-1} b$$

assuming the rows are linearly independent

$$g\left(-2(AA^T)^{-1}b\right) = -\frac{1}{4} 4\, b^T \left((AA^T)^{-1}\right)^T (AA^T)(AA^T)^{-1} b$$

$$+ 2\left(b^T (AA^T)^{-1} b\right) =$$

$$= b^T (AA^T)^{-1} b$$

Linear program (when all $f_i(x)$ are affine)

$$\min \langle c, x \rangle$$

$$\text{s.t.} \quad Ax \leq b$$

$$g(\lambda) = \inf_x \mathcal{L}(x, \lambda) = \inf_x \left(c^T x + \lambda^T (Ax - b)\right)$$

$$= \inf_x \left(c^T x + \lambda^T A x\right) + \lambda^T b =$$

$$= \inf_x \underbrace{\left(c^T + \lambda^T A\right) x}_{} + \lambda^T b$$

$c^T + \lambda^T A \neq 0$ → $-\infty$

$c^T + \lambda^T A = 0$ → 0

so the dual $\sup_{\lambda \geq 0} g(\lambda) \Rightarrow$

$$\Rightarrow \quad \max \ \lambda^T b$$

$$\text{s.t.} \quad A^T \cdot \lambda = -c$$

$$\lambda \geq 0$$

entropy maximization

$$\min \; f_0(x) = \sum_{i=1}^{n} x_i \log x_i$$

$$Ax \preceq b$$

$$1^T x = 1$$

$$L(x,\lambda,\mu) = f_0(x) + \lambda^T(Ax-b) + \mu(1^T x - 1)$$

$$= f_0(x) + \lambda^T A x + \mu 1^T x \quad - \lambda^T b - \mu$$

$$= -\left(-f_0(x) - \lambda^T A x - \mu 1^T x\right) - \lambda^T b - \mu$$

$$= -\left(\left(-\lambda^T A - \mu 1^T\right)x - f_0(x)\right) - \lambda^T b - \mu$$

$$g(\lambda,\mu) = \inf_x L(x,\lambda,\mu) =$$

$$= -\sup_x L(x,\lambda,\mu) =$$

$$= -f_0^*\left(-\lambda^T A - \mu 1^T\right) - \lambda^T b - \mu$$

$$\Rightarrow f^*(y) = e^{y-1}$$

$$\Rightarrow f^*(\mathbf{y}) = \sum_{i=1}^{n} e^{y_i - 1}$$

$$g(\lambda, \mu) = -\sum_{i=1}^{n} e^{-\lambda^T a_i - \mu - 1} - \lambda^T b - \mu$$

<u>dual program</u>

$$\max \; -e^{-(\mu+1)} \cdot \sum_{i=1}^{n} e^{-\lambda^T a_i} - \lambda^T b - \mu$$

$$\text{s.t.} \quad \lambda \geq 0$$