Your name: Rohan Chandra

1. Consider the "monotropic" program

$$\begin{aligned} \text{minimize} \quad & \|x\|_\infty \\ \text{subject to} \quad & Ax = b. \end{aligned} \tag{1}$$

(a) Write this as an unconstrained (or implicitly constrained) problem using the characteristic function of the zero vector $\chi_0(z)$. This function is zero if it's argument is zero, and infinite otherwise.

Solution:

$$\min_x ||x||_\infty + \chi_o(Ax - b)$$

(b) What is the conjugate of $f(z) = \|z\|_\infty$?

Solution: The conjugate function of $f(z) = ||z||_\infty$ is:

$$f^*(y) = ||z||_\infty{}^* = \begin{cases} 0 & ||y||_1 \leq 1 \\ \infty & otherwise \end{cases}$$

(c) What is the conjugate of $g(z) = \chi_0(z)$?

Solution: If $g(z) = \chi_0(z)$, then its conjugate is,

$$g^*(y) = \max_z y^T z - \chi_0(z) = \max_{z \in 0} y^T z$$

(d) Using the conjugate functions, write down the dual of (1).

Solution:

$$g(\lambda) = \min_x ||x||_\infty + \lambda^T(Ax - b)$$

$$\implies g(\lambda) = -b^T\lambda + \min_x(||x||_\infty + (A^T\lambda)^T x)$$

Using the conjugate norm of $||x||_\infty$, we get:

$$g(\lambda) = -b^T\lambda - ||(-A^T\lambda)||^*{}_\infty$$

$$= \begin{cases} -b^T\lambda & || - A^T\lambda||_1 \leq 1 \\ -\infty & otherwise \end{cases}$$

2. Consider the linear program

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

(a) Write the optimality conditions for this problem (i.e., the KKT system).

Solution: Let x^*, λ^*, μ^* be the optimal primal dual variables. Hence the KKT conditions are:

$$c - \lambda^* + A^T \mu^* = 0$$

$$-x^* \leq 0$$

$$Ax^* - b = 0$$

$$\lambda^* \geq 0$$

$$-\lambda^* x^* = 0$$

(b) Write the Lagrangian for this problem.

Solution:

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

$$\implies L(x, \lambda, \mu) = b^T \mu + (c + A^T \mu - \lambda)x$$

(c) Minimize out the primal variables in the Lagrangian, and write the dual formulation of this linear program.

Solution:

$$g(\lambda, \mu) = \min_x L(x, \lambda, \mu) = b^T \mu + \min_x (c + A^T \mu - \lambda)x$$

This is easily analyzed as a linear function is only bounded below when it is linear. Hence,

$$g(\lambda, \mu) = \begin{cases} -b^T \mu & c + A^T \mu - \lambda = 0 \\ -\infty & otherwise \end{cases}$$

3. Consider the problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g(x) \leq 0. \end{aligned}$$

Let x_0 be a solution to this problem, and λ_0 be the corresponding optimal Lagrange multiplier. Now, define a perturbed problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g(x) \leq \epsilon \end{aligned}$$

where ϵ is a vector. Let x_ϵ be a solution to the perturbed problem. Note, if we put large negative values in ϵ, then the constraint set gets smaller, and we expect the corresponding value of $f(x_\epsilon)$ to increase.

Prove the "sensitivity bound"

$$f(x_0) - \lambda_0^T \epsilon \leq f(x_\epsilon).$$

This bound shows that the Lagrange multipliers determine how much the objective increases as the vector ϵ becomes more negative.

Solution:

From the definition of the dual function, say for our example, $h(\lambda)$

$$h(\lambda) = \min_x f(x) + \lambda^T g(x)$$

We are given that x_ϵ is the solution to the above equation and hence, plugging $x = x_\epsilon$ in the above equation, our RHS becomes,

$$R.H.S = f(x_\epsilon) + \lambda^T g(x_\epsilon)$$

From the definition of the dual function,

$$h(\lambda) \leq f(x_\epsilon) + \lambda^T g(x_\epsilon)$$

Maximizing out the dual gives us,

$$h(\lambda_o) \leq f(x_\epsilon) + {\lambda_o}^T g(x_\epsilon)$$

Since x_ϵ is a feasible point, $g(x_e) < \epsilon$. Hence,

$$h(\lambda_o) \leq f(x_e) + {\lambda_o}^T \epsilon$$

But, from strong duality, solution of primal = solution of dual. And solution of the primal is nothing but $f(x_o)$, thus

$$f(x_o) = h(\lambda_o)$$
$$f(x_o) \leq f(x_e) + {\lambda_o}^T \epsilon$$
$$f(x_o) - {\lambda_o}^T \epsilon \leq f(x_e)$$

4. (a) Let $\text{prox}_f(x,t) = \text{argmin}_z\, f(z) + \frac{1}{2t}\|z - x\|^2$. Prove the "Moreau decomposition" identity

$$x = \text{prox}_f(x,t) + t\,\text{prox}_{f^*}(x/t, 1/t).$$

Solution: Let $\text{prox}_f(x,t) = u$. This implies

$$\partial f(u) + \frac{u-x}{t} = 0$$

$$\implies \frac{x-u}{t} \in \partial f(u)$$

Using property of sub-gradients and conjugates namely,

$$y \in \partial f(x) \iff x \in \partial f^*(y)$$

We apply this property in our proof to get

$$u \in \partial f^*(\frac{x-u}{t})$$

Now,

$$\implies \partial f^*(\frac{x-u}{t}) - u = 0$$

$$\implies \partial f^*(\frac{x-u}{t}) - (x - (x-u)) = 0$$

$$\implies \partial f^*(\frac{x-u}{t}) + t((\frac{x-u}{t}) - \frac{x}{t}) = 0$$

Reconstructing the definition of proximal operator,

$$\text{prox}_{f^*}(x/t, 1/t) = \arg\min_z f^*(z) + \frac{t}{2}||z - \frac{x}{t}||^2 = \frac{x-u}{t}$$

and

$$\text{prox}_f(x,t) = u$$

Simple addition of these two proximal operators gives

$$\text{prox}_f(x,t) + t\,\text{prox}_{f^*}(x/t, 1/t) \implies u + t(\frac{x-u}{t}) = x$$

Hence proved

(b) Using your result from part (a), prove that

$$\text{prox}_{|x|}(x,t) = x - \max\{\min\{x,t\}, -t\}$$

where $|x|$ denotes the 1-norm of x, and "min" and "max" are applied element-wise. You may use the identity $[af]^*(x) = af^*(x/a)$ where f^* denotes the conjugate of f, and a is a scalar.

Solution: By Moreau's decomposition we have,

$$\text{prox}_{||x||_1}(x,t) + \text{prox}_{||x||_\infty}(x/t, 1/t) = x$$

Using the given identity, $[af]^*(x) = af^*(x/a)$, we have

$$\text{prox}_{||x||_1}(x,t) + \text{prox}_{||tx||_\infty}(x,t) = x$$

$$\implies \text{prox}_{||x||_1}(x,t) = x - \text{prox}_{||tx||_\infty}(x,t)$$

But projecting onto the infinity ball, we know

$$\text{prox}_{||tx||_\infty}(x,t) = \begin{cases} t & x > t \\ x & |x| \leq t \\ -t & x < -t \end{cases}$$

$$\implies \text{prox}_{||tx||_\infty}(x,t) = \max\{\min\{x,t\}, -t\}$$

Hence,

$$\text{prox}_{|x|}(x,t) = x - \max\{\min\{x,t\}, -t\}$$