

1 6th of December 2018 — A. Frangioni

1.1 Duality



Do you recall?

The KKT-G condition:

$$\nabla f(x) + \sum_{i \in \mathcal{I}} \lambda_i \nabla g_i(x) + \sum_{j \in \mathcal{J}} \mu_j \nabla h_j(x) = 0$$

We would like to understand what is the function of which this is the gradient.

Definition 1.1 (Lagrangian function). *Let (P) be our minimum problem over $f : \mathbb{R}^n \rightarrow \mathbb{R}$. We term **Lagrangian function** the following:*

$$L(x, \lambda, \mu) = f(x) + \sum_{i \in \mathcal{I}} \lambda_i g_i(x) + \sum_{j \in \mathcal{J}} \mu_j h_j(x)$$

Fact 1.1. *A necessary condition for optimality is that the gradient of the Lagrangian function is 0. Formally, $\nabla L(x; \lambda, \mu) = 0$.*

A first approach to use the Lagrangian function would be taking λ and μ such that the Lagrangian is positive semidefinite.

But this is a too strict requirement, in fact we impose:

Definition 1.2 (Critical cone). *Let us assume (x, λ, μ) satisfies (KKT). We define the **critical cone** as*

$$C(x, \lambda, \mu) = \left\{ d \in \mathbb{R}^n : \begin{array}{ll} \langle \nabla g_i(x), d \rangle = 0 & i \in \mathcal{A}(x) \text{ s.t. } \lambda_i^* > 0 \\ \langle \nabla g_i(x), d \rangle \leq 0 & i \in \mathcal{A}(x) \text{ s.t. } \lambda_i^* = 0 \\ \langle \nabla h_j(x), d \rangle = 0 & j \in \mathcal{J} \end{array} \right\}$$

Theorem 1.2. *Let us assume we have a point (x, λ, μ) that satisfies the Karush-Kuhn-Tucker conditions and satisfies the linear independence of the constraints. If x is local optimum then $d^T \nabla_{xx}^2 L(x, \lambda, \mu) d \geq 0 \forall d \in C(x, \lambda, \mu)$.*

Informally, if the hypothesis holds, then the Hessian of the Lagrangian function is $\succeq 0$ on the critical cone.

Observation 1.1. *(x, λ, μ) satisfies $(KKT) \wedge \nabla_{xx}^2 L(x, \lambda, \mu) \succ 0$ on $C(x, \lambda, \mu)$ then x local optimum.*

We would like to say something more about λ and μ . Until now we considered the Lagrangian as a function of x , but what if we consider the Lagrangian in terms of λ and μ ?

1.2 Lagrangian duality

Definition 1.3 (Lagrangian relaxation). *Let us consider the Lagrangian function. We term **Lagrangian relaxation** the function where we fixed λ and μ and minimize on x :*

$$\psi(\lambda, \mu) = \min\{L(x, \lambda, \mu) : x \in \mathbb{R}^n\}$$

The relaxation leads to an unconstrained problem and we learnt how to solve one of those. This Lagrangian relaxation leads to the definition of **lagrangian dual** ψ .

Property 1.3. *The dual function has the following properties:*

- ψ is concave, but $\psi(\lambda, \mu) = -\infty$;
- ψ is non differentiable, although f , g_i and h_j are;
- Let \bar{x} be optimal in $(R_{\lambda, \mu})$, then $[\sum_{i \in \mathcal{I}} \nabla g_i(\bar{x}) + \sum_{j \in \mathcal{J}} \nabla h_j(\bar{x})] \in \partial\psi(\lambda, \mu)$
- \forall fixed $\lambda \geq 0, \mu, \bar{x} \in X$ $\psi(\lambda, \mu) = \min_x L(x, \lambda, \mu) \leq L(\bar{x}, \lambda, \mu) \leq f(\bar{x})$

$$\psi(\lambda, \mu) = \begin{pmatrix} \lambda_1 g_1(\bar{x}) \\ \vdots \\ \lambda_k g_k(\bar{x}) \\ \lambda_1 \mu_1(\bar{x}) \\ \vdots \\ \lambda_p \mu_p(\bar{x}) \end{pmatrix} \text{ is such that } \begin{pmatrix} g_1(\bar{x}) \\ \vdots \\ g_k(\bar{x}) \\ \mu_1(\bar{x}) \\ \vdots \\ \mu_p(\bar{x}) \end{pmatrix} \text{ belongs to the supergradient.}$$

Theorem 1.4 (Weak duality). \forall fixed $\lambda \in \mathbb{R}^+ \cup \{0\}, \mu \in \mathbb{R}$ such that $\psi(\lambda, \mu) \leq v(P)$ and let us take any feasible $\bar{x} \in X$ and $g(\bar{x}) \leq 0, h(\bar{x}) = 0$.

It can be proved that $\psi(\lambda, \mu) = \min_x L(x, \lambda, \mu) \leq L(\bar{x}, \lambda, \mu) \leq f(\bar{x})$.

Proof. $L(x, \lambda, \mu) = f(x) + \sum_i \lambda_i g_i(x) + \sum_j \mu_j h_j(x)$. Let us assume we have some $\bar{\lambda}, \bar{\mu}, \bar{x}$, where \bar{x} is feasible (aka $g_i(\bar{x}) \leq 0$ and $h_i(\bar{x}) = 0$ and $\bar{\lambda} \geq 0$).

The value of the dual function is

$\psi(\bar{\lambda}, \bar{\mu}) = \min_x \{f(x) + \sum_i \bar{\lambda}_i g_i(x) + \sum_j \bar{\mu}_j h_j(x)\} \leq f(\bar{x}) + \sum_i \bar{\lambda}_i g_i(\bar{x}) + \sum_j \bar{\mu}_j h_j(\bar{x})$, where the last term is cancelled, since $h_i(\bar{x}) = 0$.

On the other hand, $\sum_i \bar{\lambda}_i g_i(x) \leq 0$, which implies that $f(x) \leq 0$. □

Observation 1.2. *This theorem gives us the way to prove that a point is an optimum. Let us assume that we found $\lambda_* \leq 0, x_*$ and μ_* such that an equality holds: $\psi(\lambda_*, \mu_*) = f(x_*)$. The value of ψ gives us the information about how far we are from the optimum. Let us assume we have $\bar{\lambda}, \bar{\mu}$ and \bar{x} then $\psi(\bar{\lambda}, \bar{\mu}) \leq f(\bar{x}) \leq \psi(\bar{\lambda}, \bar{\mu}) + \varepsilon$ and this tells that we are far from the optimum of a factor ε .*

If everything in the original function was convex then the Lagrangian is convex and the computations are easy.

At this point we want to find the biggest lower bound and it may be computed since the function is concave.

$$\max\{ \psi(\lambda, \mu) : \lambda \in \mathbb{R}_+^{|\mathcal{I}|}, \mu \in \mathbb{R}^{|\mathcal{J}|} \}$$

Notice that the constraints on λ are very easy, so the problem may be considered somehow unconstrained.

We can use local methods to compute the maximum of this function.

If we are able to compute the gradient of ψ then we have the supergradients of the function f .

Is (P) equal to (D)? Yes, provided that everything is convex. In general the Lagrangian gives a lowerbound.

$$v(D) \leq v(P)$$

Example 1.1. Let us take a concave objective function $\min\{-x^2 : 0 \leq x \leq 1\}$, such that its Lagrangian function is $L(x, \lambda) = -x^2 + \lambda_1(x - 1) - \lambda_2 x$. It goes without saying that the Lagrangian function is unbounded below (upside-down parabola).

Formally, $\psi(\lambda) = \min_{x \in \mathbb{R}} L(x, \lambda) = -\infty, \forall \lambda \in \mathbb{R}^2$, which means that the Lagrangian dual is $v(D) = -\infty < v(P) = -1$.

Theorem 1.5. Let us assume that f, g and h are convex, and constraints qualification holds. If the problem has an optimum then the value of the primal and the value of the dual are the same.

$$\text{Formally, } T_X(x_*)D_X(x_*) \Rightarrow v(D) = v(P).$$

Proof. We are assuming x_* is an optimum, so the necessary conditions hold, hence we have the Karush Khun Tucker conditions, then we can find (λ^*, μ^*) that satisfy the KKT with x .

Then our claim is that (λ^*, μ^*) is an optimal solution of the dual (D) and the value of the dual is equal to the value of the primal.

x_* is a stationary point for the Lagrangian function, since it satisfies the KKT conditions then x_* is exactly the minimum, since the Lagrangian function is convex.

Hence, the value of the dual $v(D) \geq \psi(\lambda^*, \mu^*) = L(x_*, \lambda^*, \mu^*) = f(x_*) = v(P) \geq v(D)$. \square

How many solution may $\psi(\lambda, \mu)$ have? In principle many, but what if f is strongly convex, then everything in the Lagrangian is strongly convex, then the solution of ψ is unique and it is also differentiable, but typically not two times differentiable. At this point the Lagrangian dual has a single solution and the optimum of the Lagrangian dual corresponds to an optimum for f .

We only translated a minimum problem on convex functions to another minimum problem on other convex functions. We will see soon that in some cases this approach is advantageous in others it is not.

1.3 Specialized dual

1.3.1 Linead programs

$$(P) \quad \min \{ cx : Ax \geq b \}$$

Lagrangian function: $L(x, \lambda) = cx + \lambda(b - Ax) = \lambda b + (c - \lambda A)x$

Dual function:

$$\psi(\lambda) = \min_{x \in \mathbb{R}^n} L(x, \lambda) = \begin{cases} -\infty & \text{if } c - \lambda A \neq 0 \\ \lambda b & \text{if } c - \lambda A = 0 \end{cases}$$

Since we can find a minimum only on a function of the second case we have that:

$$(D) \quad \max \{ \psi(\lambda) : \lambda \geq 0 \} \equiv \max \{ \lambda b : \lambda A = c, \lambda \geq 0 \}$$

Since the lagrangian is far from having an unique optimal solution the dual does not have a unique maximum so they do not match usually.

1.3.2 Quadratic programs

Notice that the quadratic case is simpler than the linear case.

$$(P) \quad \min \left\{ \frac{1}{2} \|x\|_2^2 : Ax = b \right\} \text{ (linear least-norm solution)}$$

$L(x, \mu) = \frac{1}{2} \|x\|_2^2 + \mu(Ax - b)$, $\nabla_x L = x + \mu A = 0 \iff x = -\mu A$, which is strictly convex since x^2 is strictly convex.

$$\psi(\mu) = \min_{x \in \mathbb{R}^n} L(x, \mu) = L(-\mu A, \mu) = -\frac{1}{2} \mu^T (AA^T) \mu - \mu b$$

$$\text{Then the Lagrangian dual (D) } \max \left\{ -\frac{1}{2} \mu^T (AA^T) \mu - \mu b : \mu \in \mathbb{R}^m \right\}$$

We can generalize with quadratic functions for general problems:

Strictly convex QP: (P)

$$\min \left\{ \frac{1}{2} x^T Q x + q x : Ax \geq b \right\}, Q \succ 0 \implies (D) \quad \max \left\{ \lambda b - \frac{1}{2} v^T Q^{-1} v : \lambda A - v = q, \lambda \geq 0 \right\}$$

strong duality $\equiv v(P) = v(D)$ (almost) always holds

1.3.3 Conic program

We can do duals of things that are not quadratic programs. Sometimes we want to have non linear things to be able to draw different shapes.

Definition 1.4 (Conic program). We term **conic program** $\min \{ cx : Ax \geq_K b \}$, where $x \geq_K y \equiv x - y \in K$, where K is a convex cone.

Example 1.2. It is easy to see that the \geq constraints we saw before are a special case of conic program, where the conic is \mathbb{R}_+^n .

$$-x_1 - 2x_2 \leq 2$$

$$x_1 + x_2 \geq 1$$

$$2x_1 - x_2 \geq 0$$

Let us write $-x_1 - 2x_2 - 2 = s_0$, $x_1 + x_2 - 1 = s_1$ and $2x_1 - x_2 = s_2$, such that $s_0, s_1, s_2 \geq 0$.

We have a mapping from \mathbb{R}^n (where the variables live) to \mathbb{R}^m (where the constraints live) and $Ax - b \in K$.

Rather than writing $s_0, s_1, s_2 \geq 0$ we could write $s_0 \geq \sqrt{s_1^2 + s_2^2}$ and this would still be a conic program.

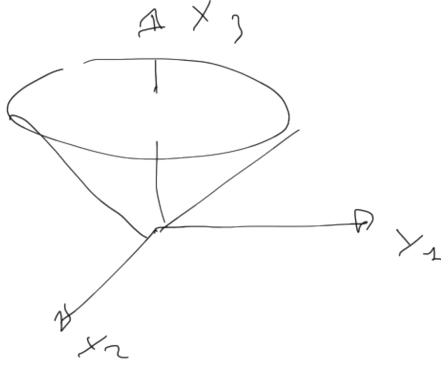


FIGURE 1.1: $x^3 \geq \sqrt{x_1^2 + x_2^2}$ is a convex cone.

The idea is to hide the non linear part in the cone, in particular in \geq_K . Let us see what happens to cones depending on the function.

There are thrww interesting cones:

- $K = \mathbb{R}_+^n \equiv \text{sign constraints} \equiv \text{Linear Program}$;
- $K = \mathbb{L} = \{x \in \mathbb{R}^n : x_n \geq \sqrt{\sum_{i=1}^{n-1} x_i^2}\} \equiv \text{Second-Order Cone (or Lorentz cone) Program, that generalizes linear programs}$;
- $K = \mathbb{S}_+ = \{A \succeq 0\} \equiv \text{"}\succeq\text{" constraints} \equiv \text{SemiDefinite Program}$.

Given the problem that looks like linear except for the cone.

Definition 1.5 (Conic dual). *Conic dual: (D) $\max\{yb : yA = c, y \geq_{K^D} 0\}$, where $K^D = \{z : \langle z, x \rangle \geq 0 \ \forall x \in K\}$ is a dual cone.*

Sometimes constraints qualification does not hold, since the conic program is not linear sometimes.

Explicit form of second order cone program (SOCP) ("explicit data" D_i, d_i, p_i, q_i) $\min\{cx : \|D_i x - d_i\|_2 \leq p_i x - q_i \ i = 1, \dots, m\}$.

It collapses to a linear program for $D_i = 0$ and $d_i = 0$.

It turns out that the dual has just this explicit form:

$$\max\left\{\sum_{i=1}^m \lambda_i d_i + \nu_i q_i : \sum_{i=1}^m \lambda_i D_i + \nu_i p_i = c, \|\lambda_i\|_2 \leq \nu_i, i = 1, \dots, m\right\}$$

where the ν_i are the dual variables of the rightmost part, while the λ_i are the dual variables of the leftmost part.

We can write the explicit form of semidefinit problems as $\min\{cx : \sum_{i=1}^n x_i A^i \succeq B\}$, where $A^i, B \in M(k, \mathbb{R})$.

There is an even mroe general form of duality, that we are going to introduce next.

1.4 Fenchel's duality

Definition 1.6 (Fenchel's conjugate). We denote **Fenchel's conjugate** of f $f^*(z) = \sup_x \{zx - f(x)\}$.

We can observe that f^* is always convex, even if f is not and it is closed if f is.

The following functions are such that f^* can be computed easily:

1. $f(x) = \frac{1}{2} \|x\|_2^2 \implies f^*(z) = \frac{1}{2} \|z\|_2^2$ (only function s.t. $f^* = f$);
2. $(\|\cdot\|_1)^*(z) = \mathbb{B}_{\infty(0,1)}(z)$, $(\|\cdot\|_\infty)^*(z) = \mathbb{B}_{1(0,1)}(z)$;
3. $f(x) = \max\{g_i x - \alpha_i \mid i \in I\} \implies f^*(z) = \min \{ \sum_{i \in I} \alpha_i \theta_i \mid \sum_{i \in I} g_i \theta_i = z, \sum_{i \in I} \theta_i = 1, \theta_i \geq 0 \mid i \in I \}$.

Fact 1.6. If we are minimizing the sum of two convex functions $(P) \min\{f(x) + g(x)\}$, this problem is the same of maximizing the sum of “almost” the two conjugates: $(D) - \min\{f^*(z) + g^*(-z)\}$.