10-725: Optimization **Fall 2013**

Lecture 13: KKT conditions

Lecturer: Ryan Tibshirani *Scribes: Guoqing Zheng, Minghao Ruan*

13.1 Continued from last lecture on duality

13.1.1 Weak Duality

Recall from last lecture, we reached the following conclusion:

$$f^* = \min_{x \in C} f(x) \geq \min_{x \in C} L(x,\, u,\, v) \geq \min_{x \in \mathbb{R}} L(x,\, u,\, v) := g(u,\, v) \tag{13.1}$$

And we denote the tight upper bound of $g(u,\, v)$ as $g^* := \max_{u>0} g(u,\, v)$.

The key insight is that the **weak duality property** always hold no matter the primal problem is convex or not, namely:

$$f^* \geq g^* \tag{13.2}$$

Also note that the dual problem is always a **convex optimization** problem (maximizing a concave function), even when the primal problem is non-convex.

By definition:

$$\begin{aligned} g(u,\, v) &= \min_{x \in \mathbb{R}^n} \left[f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j \ell_j(x) \right] && (13.3) \\ &= -\max_{x \in \mathbb{R}^n} \left[-f(x) - \sum_{i=1}^{m} u_i h_i(x) - \sum_{j=1}^{r} v_j \ell_j(x) \right] && (13.4) \end{aligned}$$

For any x, pointwise maximum is a convex function in $(u,\, v)$.

The following example illustrates this property:

$$\begin{aligned} \min_x \;\; & f(x) = x^4 - 50x^2 + 100x \\ \text{subject to} \;\; & x \geq -4.5 \end{aligned} \tag{13.5}$$

The original problem is obvious non-convex as shown in Fig. 13.1.

Though the dual function can be derived explicitly (differentiate the Lagrangian and find a closed-form solution of a cubic equation), the form of g is quite complicated but it is concave!

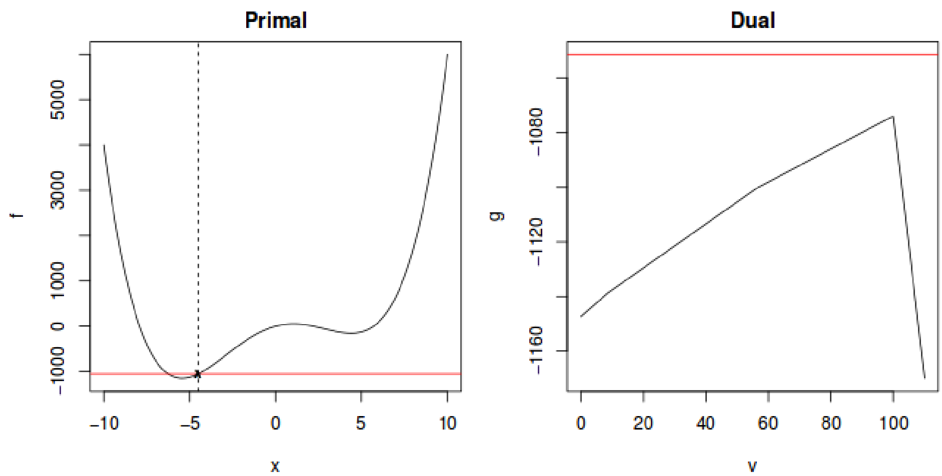

Figure 13.1: Nonconvex primal problem and its concave dual problem

13.1.2 Strong Duality

Weak duality is good but in many problems we have observed something even better:

$$f^* = g^* \tag{13.6}$$

which is called the **strong duality**. But when do we have this nice property?

Slater's Condition:

- if the primal is convex (i.e., f and $h_1, \ldots h_m$ are convex, $\ell_1, \ldots \ell_r$ are affine)
- if there exists at least one strictly feasible $x \in \mathbb{R}^n$
  (i.e., $h_1(x) < 0, \ldots h_m(x) < 0$ and $\ell_i(x) = 0, \ldots \ell_r(x) = 0$)

This is actually a weak statement and it can be further refined: need strict inequality only over h_i that are not affine.

In the case of linear programming:

- If the primal LP is feasible, then by Slater's condition strong duality holds and hence $f^* = g^*$;
- If the dual LP is feasible, then by Slater's condition strong duality holds and hence $g^* = f^*$;
- Strong duality breaks only when both primal and dual are infeasible.

13.2 Recap and Summary: Primal problem and dual problem

Primal problem:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & h_i(x) \leq 0, i = 1, ..., m \\ & \ell_j(x) = 0, j = 1, ..., r \end{aligned} \tag{13.7}$$

Lagrangian:

$$L(x,u,v) = f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j \ell_j(x) \tag{13.8}$$

Lagrange dual function:

$$g(u,v) = \min_{x\in\mathbb{R}^n} L(x,u,v) \tag{13.9}$$

Dual problem:

$$\begin{aligned} \max_{u\in\mathbb{R}^m, v\in\mathbb{R}^r} \quad & g(u,v) \\ \text{subject to} \quad & u \geq 0. \end{aligned} \tag{13.10}$$

Immediate results:

- For any feasible solution, $f(x) \geq L(x,u,v)$.
- $g(u,v)$ is always concave, even if $f(x)$ is not convex.
- Weak duality: It is always true that $f^* \geq g^*$. Hence for any (x,u,v),

  $$f(x) - f^* \leq f(x) - g(u,v) \tag{13.11}$$

  If $f(x) - g(u,v) = 0$, then x is primal optimal and (u,v) are dual optimal.
- Slater's condition: for convex primal, if there is an x such that

  $$h_1(x) < 0, ..., h_m(x) < 0 \text{ and } \ell_1(x) = 0, ..., \ell_r(x) = 0 \tag{13.12}$$

  then strong duality holds, i.e., $f^* = g^*$.

13.3 Karush-Kuhn-Tucker conditions

Theorem 13.1 *Under strong duality, x^* and u^*, v^* are primal and dual solutions if and only if the KKT conditions hold, which are:*

- **Stationarity**: $0 \in \partial f(x^*) + \sum_{i=1}^{m} u_i^* \partial h_i(x^*) + \sum_{j=1}^{r} v_j^* \partial \ell_j(x^*)$
- **Complementary slackness**: $u_i^* h_i(x^*) = 0$ *for all* i
- **Primal feasibility**: $h_i(x^*) \leq 0, \ell_j(x^*) = 0$ *for all* i, j
- **Dual feasibility**: $u_i^* \geq 0$ *for all* i.

Proof: We first prove necessity.

$$f(x^*) = g(u^*, v^*) \tag{13.13}$$
$$= \min_{x\in\mathbb{R}^n} f(x) + \sum_{i=1}^m u_i^* h_i(x) + \sum_{j=1}^r v_j^* \ell_j(x) \tag{13.14}$$
$$\leq \min_{x\in\mathbb{R}^n} f(x) \tag{13.15}$$
$$= f(x^*) \tag{13.16}$$

Hence the above inequality is actually an equality, which means

- Primal feasibility and dual feasibility obvisouly hold;
- x^* minimizes $L(x, u^*, v^*)$ over $\mathbb{R}^n$, hence the subdifferential of $L(x, u^*, v^*)$ contains 0 at $x = x^*$, which is the stationarity condition;
- $\sum_{i=1}^m u_i^* h_i(x^*) = 0$ and since $u_i^* \leq 0$ and $h_i(x) \leq 0$, hence $u_i^* h_i(x^*) = 0$ for all i, which is the complementary slackness.

Next we prove sufficiency. If there exists x^*, u^* and v^* that satisfy the KKT condition, then

$$g(u^*, v^*) = \min_{x\in\mathbb{R}^n} L(x, u^*, v^*) \tag{13.17}$$
$$= f(x^*) + \sum_{i=1}^m u_i^* h_i(x^*) + \sum_{j=1}^r v_i^* \ell_j(x^*) \qquad \text{(stationarity)} \tag{13.18}$$
$$= f(x^*) \qquad \text{(complementary slackness, dual feasibility)} \tag{13.19}$$

which means the duality gap is zero and therefore x^*, u^*, v^* are optimal solutions. ■

Warning: One may attempt to conclude that stationarity is equivalent to the following:

$$0 = \nabla f(x) + \sum_{i=1}^m u_i^* \nabla h_i(x) \sum_{j=1}^r v_j^* \nabla \ell_j(x) \tag{13.20}$$

This is only true when $f(x)$, $h_i(x)$ and $\ell_j(x)$ are convex.

Another way to formulate the problem is using the indicator function $\mathbb{I}$ and normal cone $\mathcal{N}$:

$$f(x) + \sum_{i=1} \mathbb{I}_{\{h_i(x)\leq 0\}} + \sum_{j=1} \mathbb{I}_{\{\ell_j(x)=0\}} \tag{13.21}$$
$$0 \in \partial f(x^*) + \sum_{i=1} \mathcal{N}_{\{h_i(x^*)\leq 0\}} + \sum_{j=1} \mathcal{N}_{\{\ell_j(x^*)=0\}} \tag{13.22}$$
$$\Longleftrightarrow x^* \text{is optimal} \tag{13.23}$$

13.4 Examples

13.4.1 Quadratic optimization with equality constraints

Consider for $Q \succeq 0$,

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & \frac{1}{2} x^T Q x + c^T x \\ \text{subject to } \ & Ax = 0 \end{aligned} \tag{13.24}$$

As $Q \succeq 0$, the above problem is convex. By stationarity and primal feasibility, we have x is a solution if and only if

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix} = \begin{bmatrix} -c \\ 0 \end{bmatrix} \tag{13.25}$$

for some v.

13.4.2 Side note: Newton's method on linearly constrained problem

$$\min \quad f(x) \tag{13.26}$$

$$\text{subject to} \quad Ax = b \tag{13.27}$$

Recall Newton's method updates x as follows: $x^+ = x - (\nabla^2 f(x))^{-1} \nabla f(x)$, but this update rule cannot garuantee the constraint is satisfied at each step.

Instead we can start from one x that satisfies the constraint: $Ax = b$, then we update it with:

$$x^+ = x + \Delta \qquad \text{where } A\,\Delta = 0 \tag{13.28}$$

and then minimize the following quadratic problem:

$$\frac{1}{2} \Delta^T (\nabla^2 f(x)) \Delta + \nabla f(x)^T \Delta \tag{13.29}$$

13.4.3 Water-filling

Consider

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & -\sum_{i=1}^{n} \log(\alpha_i + x_i) \\ \text{subject to } \ & x \geq 0, 1^T x = 1 \end{aligned} \tag{13.30}$$

The Lagrangian is

$$L(x, u, v) = -\sum_{i=1}^{n} \log(\alpha_i + x_i) - \sum_{i=1}^{n} u_i x_i + v(\sum_{i=1}^{n} x_i - 1)$$

Stationarity:

$$-\frac{1}{\alpha_i + x_i} - u_i + v = 0 \qquad \text{for all } i = 1, ..., n \tag{13.31}$$

Complementary slackness:

$$u_i x_i = 0 \qquad \text{for all } i = 1, ..., n \tag{13.32}$$

Primal feasibility:

$$x \geq 0, 1^T x = 1 \tag{13.33}$$

Dual feasibility:

$$u_i \geq 0 \tag{13.34}$$

Combining the above results, we get

$$\begin{aligned} v - \frac{1}{\alpha_i + x_i} &\geq 0 \\ x_i(v - \frac{1}{\alpha_i + x_i}) &= 0 \end{aligned} \tag{13.35}$$

Hence if $v < \frac{1}{\alpha_i}$, then $x_i > 0$, then $v = \frac{1}{\alpha_i + x_i}$ which is $x_i = \frac{1}{v} - \alpha_i$;

if $v \geq \frac{1}{\alpha_i}$, then $x_i = 0$. In sum, we get $x_i = \max\{0, \frac{1}{v} - \alpha_i\}$. By primal feasibility, we solve the uni-variate optimization problem

$$\sum_{i=1}^{n} \max\{0, \frac{1}{v} - \alpha_i\} = 1 \tag{13.36}$$

to get the solution to the original problem.

13.4.4 Lasso

Consider

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \tag{13.37}$$

From stationarity,

$$X^T(y - X\beta) = \lambda s \tag{13.38}$$

where $s \in \partial \|\beta\|_1$, that is

$$s_i \in \begin{cases} sign(\beta_i) & \text{if } \beta_i \neq 0 \\ [-1, 1] & \text{if } \beta_i = 0 \end{cases} \tag{13.39}$$

from which we directly get if $|X_i^T(y - X\beta)| < \lambda$, then $\beta_i = 0$.

13.4.5 Group Lasso

Consider

$$\min_{\beta=(\beta^{(1)},\beta^{(2)},...,\beta^{(G)})\in\mathbb{R}^p} \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\sum_{i=1}^{G} w_i\|\beta^{(i)}\|_2 \tag{13.40}$$

From stationarity, for $i = 1, 2, ..., G$,

$$(X^{(i)})^T(y - X\beta) = \lambda w_i s^{(i)} \tag{13.41}$$

where $s^{(i)} \in \partial\|\beta^{(i)}\|_2 = \begin{cases} \frac{\beta^{(i)}}{\|\beta^{(i)}\|_2} & \text{if } \beta^{(i)} \neq 0 \\ \{v : \|v\|_2 \leq 1\} & \text{otherwise} \end{cases}$.

- If $\|(X^{(i)})^T(y - X\beta)\|_2 < \lambda w_i \Rightarrow \beta^{(i)} = 0$.
- If $\beta^{(i)} \neq 0$,

$$\begin{aligned} &(X^{(i)})^T(y - X^{(i)}\beta^{(i)} - \sum_{j\neq i} X^{(j)}\beta^{(j)}) = \lambda w_i \frac{\beta^{(i)}}{\|\beta^{(i)}\|_2} \\ \Rightarrow &-(X^{(i)})^T X^{(i)}\beta^{(i)} + (X^{(i)})^T r^{(i)} = \lambda w_i \frac{\beta^{(i)}}{\|\beta^{(i)}\|_2} \\ \Rightarrow &\left(\lambda w_i \frac{\beta^{(i)}}{\|\beta^{(i)}\|_2} I + (X^{(i)})^T X^{(i)}\right)\beta^{(i)} = (X^{(i)})^T r^{(i)} \\ \Rightarrow &\beta^{(i)} = \left(\lambda w_i \frac{\beta^{(i)}}{\|\beta^{(i)}\|_2} I + (X^{(i)})^T X^{(i)}\right)^{-1} (X^{(i)})^T r^{(i)} \end{aligned} \tag{13.42}$$

  where $r^{(i)} = y - \sum_{j\neq i} X^{(j)}\beta^{(j)}$.

13.5 Summary

Under strong duality, we can characterize the primal solution from its dual problem.

Recall that under strong duality, the KKT conditions are necessary for optimality. Given dual solutions (u^*, v^*), any primal solution satisfies the stationarity condition:

$$0 \in \partial f(x^*) + \sum_{i=1}^{m} u_i^* \partial h_i(x^*) + \sum_{j=1}^{r} v_j^* \partial \ell_j(x^*) \tag{13.43}$$

In other words, x^* achieves the minimum in $\min_{x\in\mathbb{R}^n} L(x, u^*, v^*)$.

- In general, this reveals a characterization of primal solutions
- In particular, if this is satisfied uniquely (i.e., above problem has a unique minimizer), then the corresponding point must be the primal solution.

References

[BV04] S. Boyd and L. Vandenberghe, "Convex optimization".