CISC 371 Class 23

The Dual Formulation for Equality Constraints

Texts: [1] pp. 237–247; [2] pp. 215–229; [3] pp. 343–349

Main Concepts:
- *Dual formulation solves first for Lagrange multipliers*
- *Minimizer is deduced from the multipliers*
- *Lagrange function for dual problem*

Sample Problem, Machine Inference: How can we solve for equality constraints with the minimum number of variables?

Optimization of a quadratic objective function with a linear equality constraint, which was stated as Problem 21.20, is expressed in the *primal form.* In many contexts, it is useful to explore alternative formulations.

23.1 Considering Constrained Optimization

The basic unconstrained optimization problem, for a function $f:\mathbb{R}^n \to \mathbb{R}$ that has at least one global minimum, is

$$\vec{w}^* = \operatorname*{argmin}_{\vec{w}\in\mathbb{R}^n} f(\vec{w}) \tag{23.1}$$

Suppose that we add a set of l linearly independent equality constraints to Problem 23.1. We can write these constraints in the form of a full-rank matrix $M_{l\times n}$ and a vector $\vec{c}\in\mathbb{R}^l$, so that we further require that a solution $\vec{w}^*$ satisfies the linear equation $M\vec{w}=\vec{c}$. The problem becomes one of optimization with linear equality constraints, that we can state as

$$\begin{aligned} \vec{w}^* &= \operatorname*{argmin}_{\vec{w}\in\mathbb{R}^n} f(\vec{w}) \\ \text{with}\quad M\vec{w}^* - \vec{c} &= \vec{0} \end{aligned} \tag{23.2}$$

Consider writing Problem 23.2 in the form of Problem 23.1, but changing the domain of the argument-minimum function from $\mathbb{R}^n$ to the set of feasible points, which are all vectors $\vec{u}$ that are solutions to the linear equation $M\vec{u}=\vec{c}$. As described in the extra notes for this class, we can write the affine space of solutions as the convex set

$$\mathbb{F} = \{\vec{u} : (M\vec{u}=\vec{c})\}$$

Another form of Problem 23.2 is

$$\vec{w}^* = \underset{\vec{w} \in \mathbb{F}}{\operatorname{argmin}} f(\vec{w}) \tag{23.3}$$

There is a large class of numerical methods, called *interior-point* methods, that try to finds solutions to Problem 23.3. Rather than using one of these methods, let us try to formulate Problem 23.2 in the form of Problem 23.1 using Lagrange multipliers. The problem would then be

$$\begin{array}{rrcl} \text{For} & \mathcal{L}(\vec{w}, \vec{\mu}) & = & f(\vec{w}) + \vec{\mu}^T [M\vec{w} - \vec{c}] \\ \text{with} & \mu_j & \neq & 0 \\ \text{find} & \{\vec{w}^*, \vec{\mu}\} & = & \underset{\vec{w} \in \mathbb{R}^n,\ \vec{\mu} \in \mathbb{R}^l}{\operatorname{argmin}} \mathcal{L}(\vec{w}, \vec{\mu}) \end{array} \tag{23.4}$$

Consider applying the idea of Problem 23.3 – which is restricting the domain of the argument-minimum function – to Problem 23.4. The proposal is to simplify the domain from all $\vec{w} \in \mathbb{R}^n$ and all $\vec{\mu} \in \mathbb{R}^l$ to be only $\vec{\mu} \in \mathbb{R}^l$. To do this, we would need to "eliminate" the term $\vec{w}$ from the Lagrange function $\mathcal{L}(\vec{w}, \vec{\mu})$. That is, we want to define a new function

$$g(\vec{\mu}) \quad \text{from} \quad \mathcal{L}(\vec{w}, \vec{\mu}) \tag{23.5}$$

This seemingly minor consideration has substantial implications in optimization.

23.2 Lagrange Dual Function

There are an infinite number of ways to meet the specification in Equation 23.5. One way, which is inspired by our goal of optimization, is to define $g(\vec{\mu})$ as the function that results from minimizing the Lagrange function $\mathcal{L}(\vec{w}, \vec{\mu})$. We will use the symbol $\mathcal{L}_D$ for this new function, rather than using the symbol g.

Definition: Lagrange dual function

> For any Lagrange function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}$ that is created from a function $f : \mathbb{R}^n \to \mathbb{R}$ and l linearly independent linear equality constraints, the *Lagrange dual function* $\mathcal{L}(\vec{w}, \vec{\mu})$ is defined as
>
> $$\mathcal{L}_D(\vec{\mu}) \stackrel{\text{def}}{=} \min_{\vec{w} \in \mathbb{R}^n} \mathcal{L}(\vec{w}, \vec{\mu}) \tag{23.6}$$

A remarkable result in optimization is that, even if the underlying objective function $f(\vec{w})$ is not convex, the Lagrange dual function is *always* concave. This results – which is far from intuitive for most readers – holds also for inequality constraints, and for convex nonlinear constraints. We will state the simplest form that is needed here, with the understanding that a more general form also holds.

Theorem: Lagrange dual function is concave

For any Lagrange function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}$ that is created from a function $f : \mathbb{R}^n \to \mathbb{R}$ and l linearly independent linear equality constraints, the Lagrange dual function $\mathcal{L}(\vec{w}, \vec{\mu})$ has the property

$$\mathcal{L}_D(\vec{\mu}) \;\; \textit{is concave} \tag{23.7}$$

Proof: In most optimization texts; for example, see: Nocedal & Wright [3] pp. 343–349, Boyd & Vandenberghe [2] pp. 215–216, and Beck [1] pp. 238–247.

Observation: This theorem implies that we can write Problem 23.4 in another way. Using the dual function of Definition 23.6, we can write the minimization of the constrained objective function as

$$\min_{\vec{w}\in\mathbb{R}^n,\, \vec{\mu}\in\mathbb{R}^l} \mathcal{L}(\vec{w}, \vec{\mu}) = \min_{\vec{w}\in\mathbb{R}^n} \max_{\vec{\mu}\in\mathbb{R}^l} \mathcal{L}(\vec{w}, \vec{\mu}) \tag{23.8}$$

In Equation 23.8, the domain of the minimization function is $\mathbb{R}^n$ and the domain of the maximization function is $\mathbb{R}^l$. These are disjoint sets, which implies that the order of the functions is commutative. Thus, we can also write

$$\begin{aligned}
\min_{\vec{w}\in\mathbb{R}^n,\, \vec{\mu}\in\mathbb{R}^l} \mathcal{L}(\vec{w}, \vec{\mu}) &= \min_{\vec{w}\in\mathbb{R}^n} \max_{\vec{\mu}\in\mathbb{R}^l} \mathcal{L}(\vec{w}, \vec{\mu}) \\
&= \max_{\vec{\mu}\in\mathbb{R}^l} \min_{\vec{w}\in\mathbb{R}^n} \mathcal{L}(\vec{w}, \vec{\mu}) \\
&= \max_{\vec{\mu}\in\mathbb{R}^l} \mathcal{L}_D(\vec{\mu}) \qquad (23.9)
\end{aligned}$$

What is remarkable about Equation 23.9 is that there is another form of Problem 23.4, which is called the *primal* form of the optimization with linear equality constraints. The *dual* form of the problem is

$$\begin{array}{lrcl}
\text{For} & \mathcal{L}(\vec{w}, \vec{\mu}) & = & f(\vec{w}) + \vec{\mu}^T[M\vec{w} - \vec{c}] \\
\text{with} & \mu_j & \neq & 0 \\
\text{find} & \vec{\mu}^* & = & \operatorname*{argmax}_{\vec{\mu}\in\mathbb{R}^l} \mathcal{L}_D(\vec{\mu})
\end{array} \tag{23.10}$$

In the dual, Problem 23.10, we try to eliminate every reference to the vector $\vec{w}$; this gives us the Lagrange dual function $\mathcal{L}_D(\vec{\mu})$. We then solve Equation 23.9 to find the maximizer $\vec{\mu}^*$, from which we "reconstruct" the minimizer $\vec{w}^*$ to the primal form of the problem.

This may not be possible in general. However, it is possible in many of the basic problems that we are exploring in this course.

23.3 Squared-Norm Optimization with Linear Equality Constraints

Our most basic problem in constrained optimization has a squared-norm objective function, written as $f(\vec{w})$, and a set of linear equality constraints $M\vec{u} = \vec{c}$.

We can now re-state Problem 21.20 in terms of the set of feasible solutions $\mathbb{F} = \mathbb{A}(M, \vec{c})$ that are also solutions to the linear constraint $M\vec{u} = \vec{c}$. This re-statement will drop the constraint as a condition and put the constraint as an element of the minimization.

Problem: Primal form of a squared-norm objective function with equality constraints

$$
\begin{array}{rcl}
\vec{w}^* & = & \underset{\vec{w}\in\mathbb{F}}{\operatorname{argmin}}\, f(\vec{w}) \\
\text{where:} & & \\
f(\vec{w}) & = & \frac{1}{2}\vec{w}^T K \vec{w} \\
\mathbb{F} & = & \mathbb{A}(M, \vec{c})
\end{array}
\tag{23.11}
$$

The key to using the dual formulation to solve Problem 23.11 is that there is a simple way to write the Lagrange dual function $\mathcal{L}_D(\vec{\mu})$. The Lagrange primal function for Problem 23.11 is

$$
\begin{aligned}
\mathcal{L}(\vec{w}, \vec{\mu}) &= f(\vec{w}) + \vec{\mu}^T[M\vec{w} - \vec{c}] \\
&= \frac{1}{2}\vec{w}^T K \vec{w} + \vec{\mu}^T[M\vec{w} - \vec{c}]
\end{aligned}
\tag{23.12}
$$

Equation 23.12 is convex in the argument $\vec{w}$, so the minimum of the Lagrange primal function occurs at a stationary point. For the minimization that is stated in Definition 23.6, the corresponding minimizer is

$$
\underline{\nabla}\mathcal{L}(\vec{w}^*, \vec{\mu}^*) = \underline{0}
\tag{23.13}
$$

The gradient of Equation 23.13 is

$$
\underline{\nabla}\mathcal{L}(\vec{w}, \vec{\mu}) = \vec{w}^T K + \vec{\mu}^T M
\tag{23.14}
$$

Setting Equation 23.14 to zero, using the symbol $\vec{v}$ to be clear that we seek the interim stationary point $\vec{v}^*$, and because the symmetric positive definite matrix K is invertible and $[K^{-1}]^T = K^{-1}$, we have

$$
\begin{array}{rrcl}
 & [\vec{v}^*]^T K + \vec{\mu}^T M & = & \underline{0} \\
\rightarrow & [\vec{v}^*]^T K & = & -\vec{\mu}^T M \\
\rightarrow & [\vec{v}^*]^T & = & -\vec{\mu}^T M K^{-1} \\
\rightarrow & \vec{v}^* & = & -K^{-1} M^T \vec{\mu}
\end{array}
\tag{23.15}
$$

It is important for us to observe that the expression of $\vec{v}^*$ in Equation 23.15 gives us a solution to the minimization that we used to define the Lagrange dual function. This implies that we can write the Lagrange dual function explicitly as

$$\begin{aligned}
\mathcal{L}_D(\vec{\mu}) &= \operatorname*{argmin}_{\vec{v}\in\mathbb{R}^n} \mathcal{L}(\vec{v}, \vec{\mu}) \\
&= \mathcal{L}(\vec{v}^*, \vec{\mu}) \\
&= -\left[\frac{1}{2}\vec{\mu}^T B\vec{\mu} + \vec{c}^T\vec{\mu}\right] \qquad (23.16)
\end{aligned}$$

A detailed derivation, of a more general form of Equation 23.16, is provided in the extra notes for this class.

We can see that the bracketed term in Equation 23.16 is a quadratic function of the vector μ, which implies that the bracketed term is convex. Theorem 23.7 states that the Lagrange dual function is concave, and the negation of a convex function is concave, so this is a concrete example of how the Lagrange dual function can help us to state an optimization problem.

There is now a straightforward solution to Problem 23.11. From the equivalence of Equation 23.9, we need to find the maximizer $\vec{\mu}^*$ for Equation 23.16. This occurs at the stationary point, which is where the gradient is zero, which is

$$\begin{array}{rrcl}
 & \underline{\nabla}\mathcal{L}_D(\vec{\mu}^*) & = & \underline{0} \\
\equiv & -[\vec{\mu}^*]^T B - \vec{c}^T & = & \underline{0} \\
\equiv & [\vec{\mu}^*]^T B & = & -\vec{c}^T \\
\equiv & \vec{\mu}^* & = & B^{-1}[-\vec{c}]
\end{array} \qquad (23.17)$$

We can use Equation 23.17 to compute the optimal Lagrange multipliers $\vec{\mu}^*$, and then use Equation 23.23 to compute the optimal $\vec{w}^*$ for the Problem 23.11.

In summary, we have used the dual formulation to perform a remarkably useful transformation. The primal squared-norm problem has the form

$$\begin{bmatrix} K & M^T \\ M^T & 0 \end{bmatrix} \begin{bmatrix} \vec{w}^* \\ \vec{\mu}^* \end{bmatrix} = \begin{bmatrix} \vec{0} \\ \vec{c} \end{bmatrix}$$

This matrix is indefinite and larger than the original unconstrained problem. By using Equation 23.17, we have solved directly for $\vec{\mu}^*$ and the substituted to find $\vec{w}^*$.

This process cannot be used for general constrained optimization. It is, none the less, a useful process for many of the problems that we want to solve in this course.

23.4 Extra Notes on the Dual Formulation

Suppose that a matrix $M_{l\times n}$ is full rank and is underdetermined, so that $0 < l < n$. There are an infinite number of solutions to the linear equation $M\vec{w} = \vec{c}$. It will be convenient for us to refer to the set of solutions, and to understand more about the structure of the set.

Consider any specific, or particular, solution to the under-determined linear equation

$$\begin{array}{rcl} M_{l\times n}\,\vec{w} & = & \vec{c} \\ \text{with} & & \\ \vec{w} & \in & \mathbb{R}^n \\ \vec{c} & \in & \mathbb{R}^l \end{array} \tag{23.18}$$

with $0 < l < n$ and $M_{l\times n}$ having $\text{rank}(M) = l$.

Let $\vec{w}_0$ be any solution to Equation 23.18, so that

$$M\vec{w}_0 = \vec{c}$$

Let us write the null space of M as $\text{null}(M)$. We can now define the set of solutions to $M\vec{w} = \vec{c}$.

<u>Definition:</u> Affine space of linear solutions

For any matrix $M_{l\times n}$ for which $0 < l < n$ and $\text{rank}(M) = l$ and the null space of M is $\mathbb{V} = \{\vec{u} \in \mathbb{R}^n : M\vec{u} = \vec{0}\}$, and any vector $\vec{c} \in \mathbb{R}^l$, and any vector $\vec{w}_0$ such that $M\vec{w}_0 = \vec{c}$, the *solution space of* $M\vec{w} = \vec{c}$ is defined as

$$\mathbb{A}(M, \vec{c}) \stackrel{\text{def}}{=} \{\vec{w} \in \mathbb{R}^n : (\vec{v} \in \mathbb{V}) \rightarrow (\vec{w} = \vec{w}_0 + \vec{v})\} \tag{23.19}$$

Definition 23.19 is a complicated, technical way of saying that the solution space of $M\vec{w} = \vec{c}$ is the set of all vectors that are the sum of some particular solution $\vec{w}_0$ and any vector in $\text{null}(M)$. The dimension of $\mathbb{A}(M, \vec{c})$ is the dimension of the null space, which is $n - l$ from the rank-nullity theorem of linear algebra.

$\mathbb{A}(M, \vec{c})$ is the sum of $\vec{w}_0$ and a vector space – specifically, the null space of M, which is $\mathbb{V}$. This implies that $\mathbb{A}(M, \vec{c})$ is, by construction, a convex set.

23.4.1 Derivation for the Dual of a Quadratic Objective Function

Consider the primal form of a problem that has a quadratic objective function with linear equality constraints, which is

$$
\begin{aligned}
\vec{w}^* &= \operatorname*{argmin}_{\vec{w}\in\mathbb{F}} f(\vec{w}) \\
\text{where:} & \\
f(\vec{w}) &= \frac{1}{2}\vec{w}^T K \vec{w} + \vec{q}^T \vec{w} \\
\mathbb{F} &= \mathbb{A}(M, \vec{c})
\end{aligned}
\tag{23.20}
$$

For the minimization that is stated in Definition 23.6, the corresponding minimizer is

$$
\underline{\nabla}\mathcal{L}(\vec{v}^*, \hat{\mu}) = \underline{0} \tag{23.21}
$$

The gradient of Equation 23.21 with respect to $\vec{w}$ is

$$
\underline{\nabla}\mathcal{L}(\vec{w}, \vec{\mu}) = \vec{w}^T K + \vec{q}^T + \vec{\mu}^T M \tag{23.22}
$$

Setting Equation 23.22 to zero, using the symbol $\vec{v}$ to be clear that we seek the interim stationary point $\vec{v}^*$, and because the symmetric positive definite matrix K is invertible and $[K^{-1}]^T = K^{-1}$, we have

$$
\begin{aligned}
& [\vec{v}^*]^T K + \vec{q}^T + \vec{\mu}^T M &= \underline{0} \\
\rightarrow \quad & [\vec{v}^*]^T K &= -\vec{\mu}^T M - \vec{q}^T \\
\rightarrow \quad & [\vec{v}^*]^T &= [-\vec{\mu}^T M + \vec{q}^T] K^{-1} \\
\rightarrow \quad & \vec{v}^* &= -K^{-1}[M^T \vec{\mu} + \vec{q}]
\end{aligned}
\tag{23.23}
$$

We will use the symbol B as a temporary abbreviation, and we will also use the dot-product identity

$$
\begin{aligned}
B &\stackrel{\text{def}}{=} [M K^{-1} M^T] \\
\vec{u}^T \vec{v} &\equiv \vec{v}^T \vec{u}
\end{aligned}
$$

With these in mind, the Lagrange dual function for Problem 23.20 is

$$
\begin{aligned}
\mathcal{L}_D(\vec{\mu}) &= \operatorname*{argmin}_{\vec{v}\in\mathbb{R}^n} \mathcal{L}(\vec{v},\vec{\mu}) \\
&= \mathcal{L}(\vec{v}^*,\vec{\mu}) \\
&= \frac{1}{2}\left[\vec{v}^*\right]^T K\left[\vec{v}^*\right] + \vec{q}^T\left[\vec{v}^*\right] + \vec{\mu}^T\left[M\left[\vec{v}^*\right] - \vec{c}\right] \\
&= \frac{1}{2}[-\vec{\mu}^T M - \vec{q}^T]K^{-1}KK^{-1}[-M^T\vec{\mu} - \vec{q}] - \vec{q}^T K^{-1}[M^T\vec{\mu} + \vec{q}] \\
&\quad -\vec{\mu}^T\left[MK^{-1}[M^T\vec{\mu} + \vec{q}] - \vec{c}\right] \\
&= \frac{1}{2}[\vec{\mu}^T M]K^{-1}M^T\vec{\mu} + 2\vec{q}^T K^{-1}M^T\vec{\mu} + \vec{q}^T K^{-1}\vec{q}] - \vec{q}^T K^{-1}M^T\vec{\mu} - \vec{q}^T K^{-1}\vec{q} \\
&\quad -\vec{\mu}^T MK^{-1}M^T\vec{\mu} - \vec{\mu}^T MK^{-1}\vec{q} - \vec{c}^T\vec{\mu} \\
&= \frac{1}{2}\vec{\mu}^T B\vec{\mu} + \vec{q}^T K^{-1}M^T\vec{\mu} - \frac{1}{2}\vec{q}^T K^{-1}\vec{q} - 2\vec{q}^T K^{-1}M^T\vec{\mu} - \vec{\mu}^T B\vec{\mu} - \vec{c}^T\vec{\mu} \\
&= -\frac{1}{2}\vec{\mu}^T B\vec{\mu} - \vec{q}^T K^{-1}M^T\vec{\mu} - \vec{c}^T\vec{\mu} - \frac{1}{2}\vec{q}^T K^{-1}\vec{q} \\
&= -\left[\frac{1}{2}\vec{\mu}^T B\vec{\mu} + \vec{q}^T K^{-1}M^T\vec{\mu} + \vec{c}^T\vec{\mu}\right] - \frac{1}{2}\vec{q}^T K^{-1}\vec{q} \qquad (23.24)
\end{aligned}
$$

so

$$
\begin{aligned}
& \underline{\nabla}\mathcal{L}_D(\vec{\mu}^*) &= \underline{0} \\
\equiv\ & -[\vec{\mu}^*]^T B - \vec{q}^T K^{-1}M^T - \vec{c}^T &= \underline{0} \\
\equiv\ & -B[\vec{\mu}^*] - MK^{-1}\vec{q} - \vec{c} &= \vec{0} \qquad (23.25) \\
\equiv\ & B[\vec{\mu}^*]^T &= -MK^{-1}\vec{q} - \vec{c} \\
\equiv\ & \vec{\mu}^* &= B^{-1}[-MK^{-1}\vec{q} - \vec{c}]
\end{aligned}
$$

For a squared-norm objective function, the vector $\vec{q} = \vec{0}$ and Equation 23.25 simplifies to

$$\vec{\mu}^* = B^{-1}[-\vec{c}]$$

which is the formula presented in the main text of the notes for this class.

End of Extra Notes

References

[1] Beck A: Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB. Siam Press, 2014

[2] Boyd S, Vandenberghe L: Convex Optimization. Cambridge University Press, 2004

[3] Nocedal J, Wright S: Numerical Optimization. Springer Science & Business Media, 2006