INTUITION BEHIND PRIMAL-DUAL INTERIOR-POINT METHODS FOR LINEAR AND QUADRATIC PROGRAMMING

PETER CARBONETTO*

1. Linear programming. There are whole books devoted to the subject of linear programming. The linear programming problem crops up in all sorts of courses, from mathematical economics to linear algebra. This short monograph cannot possibly supplant any text or course. That being said, I find that most of the texts I have encountered present the key results in such a way that they are stripped of all meaning. The text leaves it up to the reader to do all the hard work in forming a connection between proof and application. In other words, most texts on linear programming are curiously devoid of intuition. There are, of course, exceptions [10].

I thought I'd take a few moments to give the basic intuition behind the interior-point approach to linear programming and, in particular, interior-point methods that take steps simultaneously in both the primal and dual variables. The 1992 article by Gonzaga [4] appears to be a well-presented and thorough review of the subject. Many of the derivations here follow from those presented in [7].

In its standard formulation, the linear program is stated as follows (see for instance [9, 12]). We are provided with a vector r of length n representing linear costs, and a collection of m constraints specified by the rows of an $m \times n$ matrix A. We'll assume as per the usual that the rows of A are linearly independent. The object is to

$$\begin{array}{ll} \text{minimize} & r^T x \\ \text{subject to} & Ax = b, \\ & x \geq 0. \end{array} \tag{1}$$

Many, many problems in the scientific world, economics, and even in ordinary, everyday life can be cast as a linear program (which is not to say that it always a good idea to do it). One interesting and silly example of a linear program is given in [11]. This problem would be easy to solve if it wasn't for the positivity constraints on x.

A linear program is just one example of a constrained optimization problem, and there are many important optimization problems that are much more complicated. However, it is best to start off with an analysis of the linear program (1) since it forms the basis for understanding more difficult optimization problems.

1.1. The Lagrange dual. We'll start off by introducing some mathematical machinery which will allow us to write down the *dual* to the linear program in standard form. It is the dual that will help us come up with a principled method to solve (1). For the optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & e(x) = 0, \\ & c(x) \leq 0, \end{array} \tag{2}$$

with n_E equality and n_C inequality constraints, the Lagrangian function is

$$L(x, y, z) = f(x) + \sum_{i=1}^{n_E} y_i e_i(x) + \sum_{i=1}^{n_C} z_i c_i(x), \tag{3}$$

*Department of Computer Science, University of British Columbia (pcarbo@cs.ubc.ca).

where y is the collection of Lagrange multipliers associated with the equality constraints, and z is the collection of multipliers corresponding to the inequality constraints. I've chosen to write down the inequality constraints so that they are always negative, instead of being positive in the linear program. I did it this way so that my derivations match those in [2]. We will, however, have to be careful that we get the correct signs on the constraints and Lagrange multipliers in all our derivations.

The Lagrange dual function is defined to be

$$q(y,z) \equiv \inf_x L(x,y,z), \tag{4}$$

where the infimum is over all x that satisfy the equality and inequality constraints in (3). The infimum is a generalization of the minimum; the infimum of a collection of points is defined to be the largest number that acts as a lower bound on this collection. We used an infimum here instead of a minimum because the Lagrangian function might not have a minimum on the feasible set, in which case the infimum is defined to be negative infinity [2]. From now on, in order to distinguish the original objective from the dual objective $q(y,z)$, we refer to $f(x)$ as the *primal* objective.

A crucial property of the Lagrange dual is that when $z \geq 0$, $q(y,z)$ is always a lower bound on the value of the primal objective $f(x)$ at the solution $x^\star$. Proving this property isn't all that hard to do; see Sec. 5.1.3 of [2]. However, this property is important enough that it is given a special name: *weak duality*.

Since the Lagrange dual is always a lower bound on the solution to the original problem, it would make sense to try and find a point (y,z) that makes $q(y,z)$ as large as possible, hence offers the best lower bound. This simple realization leads to the Lagrange dual problem:

$$\begin{array}{ll} \text{maximize} & q(y,z) \\ \text{subject to} & z \geq 0. \end{array} \tag{5}$$

Implicitly, we have an additional constraint since the bound should be non-trivial; *i.e.* the infimum in $q(y,z)$ should not be equal to negative infinity. Just as we said x is feasible if equality and inequality constraints are satisfied, we say (y,z) is *dual feasible* if $z \geq 0$ and $q(y,z) > -\infty$ are satisfied. The dual problem (5) is always a convex optimization problem, regardless whether or not the original (primal) optimization problem is convex, so it always has a unique solution $(y^\star, z^\star)$. See Sec. 5.2 of [2] for an explanation why.

These thoughts suggest that we could alternatively solve the Lagrange dual problem instead of the original problem (2). This is especially tantalizing thought since the dual problem is convex. There are a couple of reasons why this isn't necessarily a good idea: one, the dual function might not be available in closed form; two, the solution $q(y^\star, z^\star)$ to (5) is only a lower bound on $f(x^\star)$. However, under certain conditions the lower bound is the tighest possible, meaning $f(x^\star) = q(y^\star, z^\star)$. When this happens, we have *strong duality*. Strong duality is guaranteed when we have a convex optimization problem, and an appropriate constraint qualification (such as the linear independence constraint qualification or Slater's condition). The proof of strong duality under constraint qualification is not terribly straightforward. See Sec. 5.3.2 of [2] for an argument that uses the hyperplane theorem to prove strong duality under Slater's constraint qualification, and see Proposition 3.3.9 in [1] for a more succinct proof that applies the Mangasarian-Fromovitz constraint qualification.

Since the Lagrange dual acts as a lower bound to the solution for any point (y,z), it is rather obvious that the difference $\eta(x,y,z) \equiv f(x) - q(y,z)$ provides an upper bound on the difference between the value of the objective at the current point x, and the value of the objective at the solution $x^\star$. This difference is called the *duality gap*. It is very useful as a principled stopping criterion, since the duality gap tells us that our current estimate of the solution is no less accurate than the value of the duality gap. See Sec. 5.5.1 of [2] for a way to compute an upper bound on the relative accuracy (instead of the absolute accuracy) using the duality gap.

Now let's see what the Lagrange dual looks like for our linear program. The Lagrangian function of (1) is

$$L(x, y, z) = r^T x + y^T (Ax - b) - z^T x. \tag{6}$$

The minus sign appears here because the lower bounds $x \geq 0$ must be converted to upper bounds to fit the formulation (2). We will rewrite (6) as

$$L(x, y, z) = -y^T b + (A^T y + r - z)^T x. \tag{7}$$

The Lagrange dual problem for the linear program is thus

$$\begin{array}{ll} \text{maximize} & q(y, z) = -y^T b + \inf_x \{(A^T y + r - z)^T x\} \\ \text{subject to} & z \geq 0, \ q(y, z) > -\infty. \end{array} \tag{8}$$

The dual function involves an infimum over a linear function. The only time when a linear function is bounded from below (or has a minimum point) is when the slope of the function is zero. Therefore, the infimum will only be a finite number when $A^T y + r - z = 0$. (For general problems, a sufficient condition for having a bounded infimum is that the slope of the Lagrangian function vanishes; *i.e.* $\nabla_x L(x, y, z) = 0$.) This gives us an alternate way to express dual feasibility, and an alternate way to express the dual optimization problem:

$$\begin{array}{ll} \text{maximize} & -y^T b \\ \text{subject to} & A^T y + r = z, \\ & z \geq 0. \end{array} \tag{9}$$

While this formulation of the dual is perfectly correct, it differs slightly from many presentations of linear programming (e.g. [9, 12]). That's because in those presentations, a minus sign was placed in front of the equality constraint terms in the Lagrangian function (3). For the sake of consistency, we stick to the formulation of [2].

The duality gap, meanwhile, also has a simple form. Provided that the iterate is both primal feasible ($Ax = b, x \geq 0$) and dual feasible ($A^T y + r = z, z \geq 0$), the duality gap bounds the accuracy of the current estimate. It reduces to

$$\eta(x, y, z) = r^T x + b^T y = r^T x + y^T A x = x^T (A^T y + r) = x^T z. \tag{10}$$

1.2. Optimality conditions. We now have all the ingredients we need to write down the necessary and sufficient conditions for (x, y, z) to be the primal-dual solution to a linear program written in standard form: the point must be primal feasible and dual feasible, and the duality gap must vanish. Putting everything together, we have

$$\begin{array}{rll} A^T y + r = z, & z \geq 0 & \text{(dual feasibility)} \\ Ax = b, & x \geq 0 & \text{(primal feasibility)} \\ x^T z = 0 & & \text{(vanishing duality gap).} \end{array} \tag{11}$$

These equations are an instance of the famed Karush-Kuhn-Tucker (KKT) optimality conditions. Since the variables x and z must be positive, the condition that the duality gap is zero is equivalent to requiring that $x_i z_i = 0$ for all $i = 1, \ldots, n$. These nonlinear equations are called the *complementary slackness conditions*.

The KKT conditions are all linear in the primal and dual variables, with the exception of the complementary slackness conditions. A seemingly reasonable thing to do at this point would be to form a first-order Taylor series approximation to the optimality conditions (11) about the point (x, y, z), then solve the linear system to obtain the primal-dual Newton search direction $(\Delta x, \Delta y, \Delta z)$. This isn't a particularly good idea because it regularly happens that only a small step can be taken along this so-called *affine search direction* before the positivity constraints are violated. A less aggressive approach is needed: rather than attempt to eliminate the duality gap in one go, we propose the less ambitious goal of reducing the duality gap by some factor σ. The Newton step then becomes the solution to

$$\begin{bmatrix} 0 & A^T & -I \\ A & 0 & 0 \\ Z & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = - \begin{bmatrix} A^T y + r - z \\ Ax - b \\ XZ1 - \frac{\sigma \eta(x,y,z)}{n} 1, \end{bmatrix} \tag{12}$$

where X is the matrix with x along its diagonal and zeros elsewhere, Z is the matrix with z along its diagonal, and 1 is a vector of ones. The reason we divide the target duality gap by n, as you might recall, is because we've separated the original condition $x^T z = 0$ into n complementary slackness conditions. This "perturbed" Newton step forms the basis of the primal-dual interior-point method for linear programming, minus some implementation issues. Before we enter into implementation details, it is worth our while to step back for a moment and derive the Newton step from a very different point of departure.

1.3. Primal-dual interior-point methods. The classic interior-point method for constrained optimization is the barrier method. It dates back to the work of Fiacco and McCormick [3] in the 1960s. The central idea is to introduce a barrier function which penalizes points that are close to the boundary of the feasible set, obviating the need for the inequality constraints. Adopting the logarithmic penalty, the barrier subproblem for the optimization problem (2) is

$$\begin{array}{ll} \text{minimize} & f_\mu(x) \equiv f(x) - \mu \sum_{i=1}^{n_C} \log(-c_i(x)), \\ \text{subject to} & e(x) = 0. \end{array} \tag{13}$$

The quality of the barrier approximation grows as μ approaches zero. On the other hand, it is difficult to minimize the logarithmic barrier function when μ is small, as its surface can vary rapidly near the boundary of the feasible set. The optimality conditions for the equality-constrained barrier subproblem (13) are simply

$$\nabla f_\mu(x) = \nabla f(x) + \sum_{i=1}^{n_E} y_i e_i(x) - \mu \sum_{i=1}^{n_C} \frac{\nabla c_i(x)}{c_i(x)} = 0, \qquad e(x) = 0. \tag{14}$$

The exciting thing is that we can connect a point satisfying these optimality conditions to the previous discussion on duality. My claim is that for a feasible point x satisfying the conditions (14), x is guaranteed to yield a dual point (y, z) that is dual feasible, hence we can use our expression for the duality gap to compute a valid lower bound on the objective at the solution $x^\star$.

Suppose we were to define the variables $z_i \equiv -\mu/c_i(x)$. Since x is feasible, the inequality constraints are satisfied, and so z_i must be positive. Plugging z_i into the optimality conditions (14) above, we find that $\nabla_x L(x, y, z) = 0$. This result combined with the fact that each z_i is positive means that our collection of variables z must be dual feasible (because the infimum will be finite). Furthermore, if the objective is convex, then the gradient vanishes at a unique point, and the duality gap reduces to

$$\eta(x, y, z) = f(x) - q(y, z) = -\textstyle\sum_{i=1}^{n_C} z_i c_i(x) = n_C \mu. \tag{15}$$

Thus, the solution to the barrier subproblem with a logarithmic penalty function and barrier parameter μ is within $\eta = n_C\mu$ of the solution to the original problem. What we have discovered here is a principled way to derive the perturbed primal-dual Newton search direction (12): construct a sequence of barrier subproblems with an adaptive choice of barrier parameter μ that depends on the current duality gap at (x, y, z), and a desired reduction σ. Thus, we can replace the term $\sigma\eta(x, y, z)/n$ in (12) with μ. This is the basic formula behind the primal-dual interior-point method.

It is generally fair to assume that the current iterate is both primal and dual feasible, in which case the linearized primal-dual system is given by

$$\begin{bmatrix} 0 & A^T & -I \\ A & 0 & 0 \\ Z & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = - \begin{bmatrix} 0 \\ 0 \\ XZ1 - \mu 1, \end{bmatrix} \tag{16}$$

It is easy to show that since the Newton step satisfies $A\Delta x = 0$, and by primal feasibility $Ax = b$, the new point $x + \alpha\Delta x$ is primal feasible for any step size α. Similarly, it is easy to show that the new point $(y + \alpha\Delta y, z + \alpha\Delta z)$ will be dual feasible for every choice of step length α. When the iterates are not primal and dual feasible, we no longer possess such guarantees. Under this scenario, setting the barrier parameter μ according to the duality gap is purely heuristic, as the quantity $\eta(x, y, z)$ is no longer the the difference between the values of the primal and dual objectives.

One important issue we haven't touched upon is the choice of centering parameter. The choice for σ in Mehrotra's predictor-corrector algorithm [5, 6] is

$$\sigma = \left(\frac{\eta_{\mathrm{aff}}(x, y, z)}{\eta(x, y, z)} \right)^3, \tag{17}$$

where $\eta(x, y, z)$ is the duality gap at the current point (x, y, z), and $\eta_{\mathrm{aff}}(x, y, z)$ is the duality gap that would be achieved if we were to follow the largest feasible step along the affine scaling direction; *i.e.* using the search direction that is the solution to (16) with $\mu = 0$. Let me give a brief rationale behind this choice of centering parameter. If a step along the affine scaling direction is able to make a large reduction in the duality gap, then we might as well follow it closely and make σ small. When $\sigma = 0$, we get the pure, unperturbed Newton or "affine scaling" step. On the other hand, if the affine scaling direction makes very little progress, we should emphasize a centering step, since it will set the stage for a larger reduction in the next iteration. At the other extreme, when $\sigma = 1$, the Newton direction defines a step that solely tries to center the iterate so that the pairwise products $x_i z_i$ are identical to the average of the current duality gap, and makes no attempt to reduce the duality gap.

2. Quadratic programming. Let's now look at the convex quadratic program with inequality constraints stated as follows:

$$\begin{array}{ll} \text{minimize} & f(x) \equiv \frac{1}{2} x^T H x + r^T x \\ \text{subject to} & Ax \leq b. \end{array} \tag{18}$$

where the $m \times n$ matrix A and the vector b of length m specify the linear inequality constraints. Provided the $n \times n$ symmetric matrix H is positive-definite, the objective is convex. Let's apply the analysis above to this quadratic program. This analysis is along the lines of what is presented in [8]. Since the direction of the inequality matches (2), we won't have to worry so much about getting the signs correct.

The Lagrangian function for this problem is

$$L(x, z) = f(x) + z^T c(x) \tag{19}$$

where $c(x) = Ax - b$ is the vector-valued inequality constraint function. The Lagrange dual function is, as before, defined to be $q(z) = \inf_x L(x, z)$. Since we've assumed the objective is convex, the minimum of the Lagrangian function is always bounded, so only $z \geq 0$ is the only condition for dual feasibility for the convex quadratic program. The minimum value of the Lagrangian is achieved precisely when its gradient vanishes, and it is specified by a simple expression:

$$\nabla_x L(x, z) = Hx + q + A^T z = 0. \tag{20}$$

Therefore, we can write the dual to the qudratic program (18) as follows:

$$\begin{array}{ll} \text{maximize} & L(x, z) \\ \text{subject to} & Hx + q + A^T z = 0, \\ & z \geq 0. \end{array} \tag{21}$$

The duality gap also has a very simple expression. It is

$$\eta(x, z) = f(x) - q(z) = -c(x)^T z, \tag{22}$$

which is very much similar to the duality gap $x^T z$ we had earlier for the linear program. There is an important subtlety that needs to be mentioned here: the condition (20) was not used in the derivation of this expression for the duality gap. However, the expression (22) is meaningless when (20) is not satisfied, because then it will no longer be the difference between the primal and dual objectives (the dual objective is defined only at the maximum of the Lagrangian), and we will no longer possess a lower bound on the value of the objective at the solution. By combining these conditions, we see that the optimality conditions for the quadratic program are

$$\begin{array}{rl} Hx + q + A^T z = 0 & \quad \text{(infimum condition)} \\ c(x)^T z = 0 & \quad \text{(vanishing duality gap)} \\ Ax \geq b, z \geq 0 & \quad \text{(primal and dual feasibility).} \end{array} \tag{23}$$

These are our KKT conditions for the convex quadratic program (18). The first condition is often erroneously called the dual feasibility condition, but that would be incorrect as it is not needed to ensure that the value of $q(z)$ is finite. Therefore, I have chosen to call it the "infimum condition" instead.

As before, we can rewrite the second equation as a collection of complementary slackness conditions $CZ1 = 0$, where Z is the matrix with the Lagrange multipliers z along its diagonal, and C is the $m \times m$ matrix with the inequality constraint function responses $c(x) = Ax - b$ along its diagonal. We can derive a perturbed version of these optimality conditions by constructing the logarithmic barrier. From the previous discussion, we know that this amounts to changing the complementary slackness conditions from $CZ1 = 0$ to $CZ1 = -\mu 1$, where the barrier parameter μ is equal to

$$\mu = \sigma \eta(x, z)/m = -\sigma c(x)^T z/m. \tag{24}$$

A minus sign is needed in front of the barrier parameter because the barrier parameter will always be positive but $c(x)$ at a feasible point x is always negative.

Forming a first-order Taylor-series expansion of the optimality conditions (23) perturbed by μ, we get the linear system

$$\begin{bmatrix} H & A^T \\ ZA & C \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \end{bmatrix} = - \begin{bmatrix} Hx + q + A^T z \\ CZ1 + \mu 1 \end{bmatrix}, \tag{25}$$

and its solution $(\Delta x, \Delta z)$ is the primal-dual interior-point Newton search direction. Of course, any step $(x + \alpha\Delta x, z + \alpha\Delta z)$ that we take should be checked to make sure that it remains within the feasible set defined by the inequalities $Ax \leq b$ and $z \geq 0$. It is easy to show that if the current point (x, z) satisfies the infimum condition, then any step following the Newton search direction will continue to satisfy this condition.

REFERENCES

[1] D. P. Bertsekas, *Nonlinear programming*, Athena Scientific, 1999.

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

[3] A. V. Fiacco and G. P. McCormick, *Nonlinear programming: sequential unconstrained minimization techniques*, John Wiley and Sons, 1968.

[4] C. C. Gonzaga, *Path-following methods for linear programming*, SIAM Review, 34 (1992), pp. 167–224.

[5] N. Gould, D. Orban, and P. Toint, *Numerical methods for large-scale nonlinear optimization*, Acta Numerica, 14 (2005), pp. 299–361.

[6] I. J. Lustig, R. E. Marsten, and D. F. Shanno, *On implementing Mehrotra's predictor-corrector interior-point method for linear programming*, SIAM Journal on Optimization, 2 (1992), pp. 435–449.

[7] R. D. C. Monteiro and I. Adler, *Interior path following primal-dual algorithms, part I: linear programming*, Mathematical Programming, 44 (1989), pp. 27–41.

[8] ———, *Interior path following primal-dual algorithms, part II: convex quadratic programming*, Mathematical Programming, 44 (1989), pp. 43–66.

[9] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, 2nd ed., 2006.

[10] G. Strang, *Linear Algebra and its Applications*, Academic Press, second ed., 1980.

[11] R. V. Vohra, *Advanced Mathematical Economics*, Routledge, 2005.

[12] M. H. Wright, *Interior methods for constrained optimization*, Acta Numerica, (1992), pp. 341–407.