Primal-Dual Interior-Point Methods

Javier Peña
Convex Optimization 10-725/36-725

Last time: duality revisited

Consider the problem

$$\begin{array}{rl} \min\limits_{x} & f(x) \\ \text{subject to} & Ax = b \\ & h(x) \leq 0 \end{array}$$

Lagrangian

$$L(x, u, v) = f(x) + u^{\mathsf{T}} h(x) + v^{\mathsf{T}} (Ax - b)$$

We can rewrite the primal problem as

$$\min_{x} \max_{\substack{u,v \\ u \geq 0}} L(x, u, v)$$

Dual problem

$$\max_{\substack{u,v \\ u \geq 0}} \min_{x} L(x, u, v)$$

Optimality conditions

Assume $f, h_1, \dots, h_m$ are convex and differentiable. Assume also that strong duality holds.

KKT optimality conditions for primal and dual

$$\begin{aligned}
\nabla f(x) + \nabla h(x) u + A^\mathsf{T} v &= 0 \\
U h(x) &= 0 \\
Ax &= b \\
u, -h(x) &\geq 0.
\end{aligned}$$

Here $U = \mathrm{Diag}(u)$, $\nabla h(x) = \begin{bmatrix} \nabla h_1(x) & \cdots & \nabla h_m(x) \end{bmatrix}$

Central path equations

Barrier problem

$$\min_x \quad f(x) + \tau\phi(x)$$
$$Ax = b$$

where

$$\phi(x) = -\sum_{i=1}^m \log(-h_i(x)).$$

Optimality conditions for barrier problem (and its dual)

$$\begin{aligned} \nabla f(x) + \nabla h(x)u + A^\mathsf{T} v &= 0 \\ Uh(x) &= -\tau 1 \\ Ax &= b \\ u, -h(x) &> 0. \end{aligned}$$

Useful fact: solution $(x(\tau), u(\tau), v(\tau))$ has duality gap

$$f(x(\tau)) - \min_x L(x, u(\tau), v(\tau)) = m\tau.$$

Outline

Today:

- Primal-dual interior-point method
- Special case: linear programming
- Extension to semidefinite programming

Barrier method versus primal-dual method

Like the barrier method, primal-dual interior-point methods aim to compute (approximately) points on the central path.

Main differences between primal-dual and barrier methods:

- Primal-dual interior-point methods usually take one Newton step per iteration (no additional loop for the centering step).
- Primal-dual interior-point methods are not necessarily feasible.
- Primal-dual interior-point methods are typically more efficient. Under suitable conditions they have better than linear convergence.

Central path equations and Newton step

Central path equations:

$$\begin{aligned} \nabla f(x) + \nabla h(x) u + A^\mathsf{T} v &= 0 \\ U h(x) + \tau \mathbf{1} &= 0 \\ Ax - b &= 0 \\ u, -h(x) &> 0. \end{aligned}$$

Newton step:

$$\begin{bmatrix} \nabla^2 f(x) + \sum_i u_i \nabla^2 h_i(x) & \nabla h(x) & A^\mathsf{T} \\ U \nabla h(x)^\mathsf{T} & H(x) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta u \\ \Delta v \end{bmatrix} = -r(x, u, v)$$

where

$$r(x, u, v) := \begin{bmatrix} \nabla f(x) + \nabla h(x) u + A^\mathsf{T} v \\ U h(x) + \tau \mathbf{1} \\ Ax - b \end{bmatrix}, \; H(x) = \mathsf{Diag}(h(x))$$

Surrogate duality gap, residuals

Define the dual, central, and primal residuals at current (x, u, v) as

$$\begin{aligned} r_{\text{dual}} &= \nabla f(x) + \nabla h(x) u + A^\mathsf{T} v \\ r_{\text{cent}} &= U h(x) + \tau \mathbf{1} \\ r_{\text{prim}} &= Ax - b \end{aligned}$$

Given x, u with $h(x) \le 0,\ u \ge 0$, the surrogate duality gap is

$$-h(x)^\mathsf{T} u$$

This is a true duality gap when $r_{\text{dual}} = 0$ and $r_{\text{prim}} = 0$.

Observe that (x, u, v) is on the central path if and only if $u > 0, h(x) < 0$ and

$$r(x, u, v) = 0 \ \text{ for } \ \tau = -\frac{h(x)^\mathsf{T} u}{m}.$$

Given x, u such that $h(x) \le 0,\ u \ge 0$, define $\tau(x,u) := -\frac{h(x)^\top u}{m}$.

Primal-Dual Algorithm

1. Choose $\sigma \in (0,1)$
2. Choose (x^0, u^0, v^0) such that $h(x^0) < 0,\ u^0 > 0$
3. For $k = 0, 1, \ldots$
   - Compute Newton step for

$$(x,u,v) = (x^k, u^k, v^k),\ \tau := \sigma\tau(x^k, u^k)$$

   - Choose steplength θ_k and set

$$(x^{k+1}, u^{k+1}, v^{k+1}) := (x^k, u^k, v^k) + \theta_k(\Delta x, \Delta u, \Delta v)$$

Parallel notation in the barrier method:

$$\tau = \frac{1}{t}, \quad \sigma = \frac{1}{\mu}.$$

Backtracking line search

At each step, we need to find θ and set

$$x^+ = x + \theta \Delta x, \quad u^+ = u + \theta \Delta u, \quad v^+ = v + \theta \Delta v.$$

Two main goals:

- Maintain $h(x) < 0, \ u > 0$
- Reduce $\|r(x, u, v)\|$

Use a multi-stage backtracking line search for this purpose: start with largest step size $\theta_{\max} \le 1$ that makes $u + \theta \Delta u \ge 0$:

$$\theta_{\max} = \min \left\{ 1, \ \min\{-u_i / \Delta u_i : \Delta u_i < 0\} \right\}$$

Then, with parameters $\alpha, \beta \in (0, 1)$, we set $\theta = 0.99 \theta_{\max}$, and

- Update $\theta = \beta \theta$, until $h_i(x^+) < 0$, $i = 1, \dots m$
- Update $\theta = \beta \theta$, until $\|r(x^+, u^+, v^+)\| \le (1 - \alpha \theta) \|r(x, u, v)\|$

Special case: linear programming

Consider

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

for $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$.

Dual:

$$\begin{aligned} \max_{y,s} \quad & b^{\mathsf{T}} y \\ \text{subject to} \quad & A^{\mathsf{T}} y + s = c \\ & s \geq 0 \end{aligned}$$

Some history

- Dantzig (1940s): the simplex method, still today is one of the most well-known/well-studied algorithms for LPs
- Klee and Minty (1972): pathological LP with n variables and $2n$ constraints, simplex method takes 2^n iterations to solve
- Khachiyan (1979): polynomial-time algorithm for LPs, based on ellipsoid method of Nemirovski and Yudin (1976). Strong in theory, weak in practice
- Karmarkar (1984): interior-point polynomial-time method for LPs. Fairly efficient (US Patent 4,744,026, expired in 2006)
- Renegar (1988): Newton-based interior-point algorithm for LP. Best known theoretical complexity until very recent work by Lee-Sidford.
- Modern state-of-the-art LP solvers typically use both simplex and interior-point methods

Optimality conditions and central path equations

Optimality conditions for previous primal-dual pair of linear programs

$$\begin{aligned} A^{\mathsf{T}}y + s &= c \\ Ax &= b \\ XS\mathbf{1} &= 0 \\ x, s &\geq 0 \end{aligned}$$

Central path equations

$$\begin{aligned} A^{\mathsf{T}}y + s &= c \\ Ax &= b \\ XS\mathbf{1} &= \tau\mathbf{1} \\ x, s &> 0 \end{aligned}$$

Primal-dual method versus barrier method

Newton equations for primal-dual method

$$\begin{bmatrix} 0 & A^\mathsf{T} & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = - \begin{bmatrix} A^\mathsf{T} y + s - c \\ Ax - b \\ XS\mathbf{1} - \tau\mathbf{1} \end{bmatrix}$$

Simple observation:

$$XS\mathbf{1} = \tau\mathbf{1} \Leftrightarrow s = \tau X^{-1}\mathbf{1} \Leftrightarrow x = \tau S^{-1}\mathbf{1}.$$

Hence can eliminate either s or x to get optimality conditions for either *primal* or *dual* barrier problems.

Newton steps for barrier problems

Primal and dual central path equations

$$\begin{aligned} A^{\mathsf T} y + \tau X^{-1}\mathbf{1} &= c \\ Ax &= b \\ x &> 0 \end{aligned} \qquad\qquad \begin{aligned} A^{\mathsf T} y + s &= c \\ \tau A S^{-1}\mathbf{1} &= b \\ s &> 0 \end{aligned}$$

Primal Newton step

$$\begin{bmatrix} \tau X^{-2} & A^{\mathsf T} \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = -\begin{bmatrix} A^{\mathsf T} y + \tau X^{-1}\mathbf{1} - c \\ Ax - b \end{bmatrix}$$

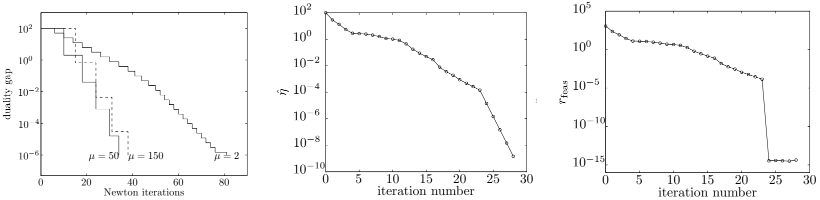
Dual Newton step

$$\begin{bmatrix} A^{\mathsf T} & I \\ 0 & \tau A S^{-2} \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta s \end{bmatrix} = -\begin{bmatrix} A^{\mathsf T} y + s - c \\ \tau A S^{-1}\mathbf{1} - b \end{bmatrix}$$

Example: barrier versus primal-dual

Example from B & V 11.3.2 and 11.7.4: standard LP with $n = 50$ variables and $m = 100$ equality constraints

Barrier method uses various values of μ, primal-dual method uses $\mu = 10$. Both use $\alpha = 0.01$, $\beta = 0.5$



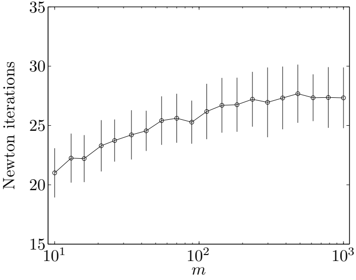
Barrier duality gap

Primal-dual surrogate duality gap

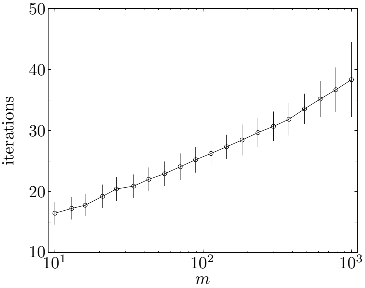
Primal-dual feasibility gap, $r_{\text{feas}} = (\|r_{\text{prim}}\|_2^2 + \|r_{\text{dual}}\|_2^2)^{1/2}$

Can see that primal-dual is faster to converge to high accuracy

Now a sequence of problems with $n = 2m$, and n growing. Barrier method uses $\mu = 100$, runs just two outer loops (decreases duality gap by 10^4); primal-dual method uses $\mu = 10$, stops when duality gap and feasibility gap are at most 10^{-8}



Barrier method



Primal-dual method

Primal-dual method require only slightly more iterations, despite the fact that they it is producing higher accuracy solutions

Interior-point methods for semidefinite programming

Primal

$$\begin{array}{ll} \min\limits_{X} & C \bullet X \\ \text{subject to} & A_i \bullet X = b_i,\ i = 1, \ldots, m \\ & X \succeq 0. \end{array}$$

Dual

$$\begin{array}{ll} \max\limits_{y} & b^{\mathsf{T}} y \\ \text{subject to} & \sum\limits_{i=1}^{m} y_i A_i + S = C \\ & S \succeq 0. \end{array}$$

Recall trace inner product in $\mathbb{S}^n$

$$X \bullet S = \mathsf{trace}(XS).$$

Strong duality holds and primal and dual attained if both are strictly feasible.

Optimality conditions for semidefinite programming

Primal and dual problems

$$\begin{array}{rl} \min\limits_{X} & C \bullet X \\ \text{subject to} & \mathcal{A}(X) = b \\ & X \succeq 0 \end{array} \qquad \begin{array}{rl} \max\limits_{y,S} & b^{\mathsf{T}} y \\ \text{subject to} & \mathcal{A}^*(y) + S = C \\ & S \succeq 0 \end{array}$$

Here $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ linear map.

Assume also that strong duality holds. Then $X^\star$ and $(y^\star, S^\star)$ are respectively primal and dual optimal solutions if and only if $(X^\star, y^\star, S^\star)$ solves

$$\begin{aligned} \mathcal{A}^*(y) + S &= C \\ \mathcal{A}(X) &= b \\ XS &= 0 \\ X, S &\succeq 0. \end{aligned}$$

Central path for semidefinite programming

Primal barrier problem

$$\begin{array}{rl} \min\limits_{X} & C \bullet X - \tau \log(\det(X)) \\ \text{subject to} & \mathcal{A}(X) = b \end{array}$$

Dual barrier problem

$$\begin{array}{rl} \max\limits_{y,S} & b^{\mathsf{T}} y + \tau \log(\det(S)) \\ \text{subject to} & \mathcal{A}^*(y) + S = C \end{array}$$

Optimality conditions for both

$$\begin{aligned} \mathcal{A}^*(y) + S &= C \\ \mathcal{A}(X) &= b \\ XS &= \tau I \\ X, S &\succ 0. \end{aligned}$$

Newton step

Primal central path equations

$$\begin{aligned} \mathcal{A}^*(y) + \tau X^{-1} &= C \\ \mathcal{A}(X) &= b \\ X &\succ 0 \end{aligned}$$

Newton equations

$$\begin{aligned} \tau X^{-1} \Delta X X^{-1} + \mathcal{A}^*(\Delta y) &= -(\mathcal{A}^*(y) + \tau X^{-1} - C) \\ \mathcal{A}(\Delta X) &= -(\mathcal{A}(X) - b) \end{aligned}$$

Similar dual central path and Newton equations involving (y, S).

Primal-dual Newton step

Recall central path equations

$$\begin{bmatrix} \mathcal{A}^*(y) + S - C \\ \mathcal{A}(X) - b \\ XS \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \tau I \end{bmatrix}, \ X, S \succ 0.$$

"Natural" Newton step:

$$\begin{bmatrix} 0 & \mathcal{A}^* & I \\ \mathcal{A} & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta X \\ \Delta y \\ \Delta S \end{bmatrix} = - \begin{bmatrix} \mathcal{A}^*(y) + S - C \\ \mathcal{A}(X) - b \\ XS - \tau I \end{bmatrix}.$$

But we run into issues of symmetry...

Nesterov-Todd direction

We want to linearize

$$XS - \tau I = 0.$$

Primal linearization:

$$S - \tau X^{-1} = 0 \rightsquigarrow \tau X^{-1} \Delta X X^{-1} + \Delta S = \tau X^{-1} - S.$$

Dual linearization:

$$X - \tau S^{-1} = 0 \rightsquigarrow \Delta X + \tau S^{-1} \Delta S S^{-1} = \tau S^{-1} - X.$$

Nesterov-Todd direction

Proper primal-dual linearization: average of previous two

$$W^{-1}\Delta X W^{-1} + \Delta S = \tau X^{-1} - S$$

or equivalently

$$\Delta X + W \Delta S W = \tau S^{-1} - X$$

provided

$$WSW = X.$$

Achieve the above by taking W as the geometric mean of X, S:

$$\begin{aligned} W &= S^{-1/2}(S^{1/2} X S^{1/2})^{1/2} S^{-1/2} \\ &= X^{1/2}(X^{1/2} S X^{1/2})^{-1/2} X^{1/2} \end{aligned}$$

Given $X, S \succeq 0$, define $\tau(X, S) := \frac{X \bullet S}{n}$.

Primal-Dual Algorithm for Semidefinite Programming

1. Choose $\sigma \in (0, 1)$
2. Choose (X^0, y^0, S^0) such that $X^0, S^0 \succ 0$
3. For $k = 0, 1, \ldots$
   - Compute Nesterov-Todd direction for

$$(X, y, S) = (X^k, y^k, S^k), \; \tau := \sigma\tau(X^k, S^k)$$

   - Choose steplength θ_k and set

$$(X^{k+1}, y^{k+1}, S^{k+1}) := (X^k, y^k, S^k) + \theta_k(\Delta X, \Delta y, \Delta S)$$

References and further reading

- S. Boyd and L. Vandenberghe (2004), "Convex optimization," Chapter 11
- S. Wright (1997), "Primal-dual interior-point methods," Chapters 5 and 6
- J. Renegar (2001), "A mathematical view of interior-point methods"
- Y. Nesterov and M. Todd (1998), "Primal-dual interior-point methods for self-scaled cones." SIAM J. Optim.