Programmation mathématique
ROP 771
Notes on predictor-corrector algorithms

Jean-Pierre Dussault

April 11, 2017

Abstract

My personnal vision of predictor-corrector interior point methods.

My background in non-linear optimization (NLO) biases my view on predictor-corrector algorithms in linear optimization (LO). Such algorithms (in LO) are frequently derived from log-barrier algorithms in NLO. Consider the NLO problem

$$\begin{aligned} &\min f_0(x) \\ \text{subject to } &f_{\leqslant}(x) \leqslant 0, \\ &f_{=}(x) = 0. \end{aligned}$$

1 Linear optimization

Consider the (standard) linear optimization instance

$$\begin{aligned} &\min cx \\ \text{subject to } &x \geqslant 0 \\ &Ax - b = 0, \end{aligned} \tag{1}$$

where c is a line vector.

1.1 Strong duality and optimality conditions

A consequence of the strong duality theorem in linear optimization is that for any optimal solution x^*, there exist y^* and $s^* \geqslant 0$ such that

$$\begin{aligned} Ax^* - b &= 0 \\ y^*A + s^* - c &= 0 \\ s_i^* x_i^* &= 0 \\ s^*, x^* &\geqslant 0. \end{aligned}$$

The last componentwize equation is best rewritten as $S^*X^*\mathbb{1} = S^*x^* = 0$ where $S^* = diag(s^*)$ and $X^* = diag(x^*)$. Thus any (primal-dual) solution (x, y, s) of the linear optimization problem satisfies the following (bi-linear) equations for $x, s \geqslant 0$:

$$\Phi(x, y, s) = \left\{ \begin{array}{l} Ax - b \\ yA + s - c \\ Sx \end{array} \right\} = 0. \tag{2}$$

The ellipsoïd method decides in polynomial time whether the system possesses a solution or not using the affine version of the duality gap $Sx = cx - yb$ in the last equation together with the bounds $x \geqslant 0$ and $s \geqslant 0$. As such, the ellipsoïd method accept a polyhedral set described by linear equations and inequalities.

Interior point methods use the bilinear version of the duality gap and assume that $x, s > 0$ is available, i.e. x and s are interior points of their respective non-negative orthant domain.

1.2 From strong duality to Newton iterations

Since equation (2) is a non-linear, actually a bi-linear equation, coupled with the bounds on x and s, its solution is not expressed as an analytical formula. As a non-linear equation, one could consider an iterative solution using Newton's method.

$$\nabla\Phi(x, y, s) = \begin{bmatrix} A & 0 & 0 \\ 0 & A^t & I \\ S & 0 & X \end{bmatrix}, \tag{3}$$

so that given some $w = (x, y, s)$, an improved solution would be given by

$$(x^+, y^+, s^+) = (x, y, s) - \nabla\Phi(x, y, s)^{-1}\Phi(x, y, s). \tag{4}$$

The very nature of the bi-linear equations $Sx = 0$ makes this Newton iteration process doubtful, but it is very efficient when it happens to converge.

1.3 From the logarithmic barrier to primal-dual equations

Returning to the particular instance (1), we may treat the non-negativity constraints with a log barrier to get

$$\begin{array}{ll} \min\limits_{x>0} & cx - \rho \sum \log(x_i) \\ \text{subject to} & \\ & Ax - b = 0, \end{array} \tag{5}$$

for which the Lagrange optimality conditions are

$$\left. \begin{array}{r} Ax - b \\ yA + \rho X^{-1}\mathbb{1} - c \end{array} \right\} = 0 \tag{6}$$

and substituting $s = \rho X^{-1}\mathbb{1}$, compare to (2), (still using Φ but with a fourth parameter ρ)

$$\Phi(x, y, s, \rho) = \left\{ \begin{array}{r} Ax - b \\ yA + s - c \\ SX\mathbb{1} - \rho\mathbb{1} \end{array} \right\} = 0. \tag{7}$$

The bi-linear equation is no longer rooted to 0 until $\rho \searrow 0$ reaches 0, so that an homotopy like strategy appears promising.

Thus, we have equations (7) that define $w(\rho) = (x(\rho), y(\rho), x(\rho))$ a point on the so-called central trajectory. The approach is to track approximate such points toward $\rho = 0$. Therefore, let us assume we have a $w = (x, y, s)$ such that $x > 0, s > 0$ and

$$\begin{pmatrix} Ax - b \\ yA + s - c \\ SXe - \rho\mathbb{1} \end{pmatrix} = \begin{pmatrix} r_P \\ r_D \\ r_C - \rho\mathbb{1} \end{pmatrix}, \tag{8}$$

where r_P is the primal, r_D the dual and r_C the complementarity residuals. Algorithms will be designed to

1. drive the residuals r to zero
2. drive the parameter ρ to zero.

For fixed ρ, Newton method's may be applied to drive the residual vector r to zero. Once this residual is small enough, a pseudo Newton method may be used to recover acceptable residual for a reduced value of ρ. Let us now describe the details of those steps.

1.4 General predictor-corrector

First, let us rewrite the system (7) in an abstract way as

$$\Phi(w, \rho) = \Phi(w) + \rho v \tag{9}$$

For a current iterate w, $r = \Phi(w)$. We then may interpret the current iterate as a function $w(\rho, r)$. Newton's method to solve $\Phi(w, \rho) = 0$ for some fixed ρ is written

$$w(\rho, 0) \approx w(\rho, r) - (\nabla\Phi(w))^{-1}\,\Phi(w, \rho) = w - (\nabla\Phi(w))^{-1}\,(r + \rho v) = w(\rho, r) + (0 - r)\dot{w}_r. \tag{10}$$

The solution $w(\rho, 0)$ to equation $\Phi(w, \rho) = 0$ for fixed ρ is a point on the central trajectory. The implicit function theorem yields that $\nabla\Phi(w)\dot{w} + v = 0$ so that an estimate of $w(0, r)$ for fixed r is given by

$$w(0, r) \approx w(\rho, r) + (0 - \rho)\dot{w}_\rho = w(\rho, r) - (\nabla\Phi(w))^{-1}(-\rho v). \tag{11}$$

We see that a Newton's iteration and the extrapolation both use $\nabla\Phi(w)$, which have us interpret the Newton iteration as an extrapolation from r to 0 of $w(\rho, r)$, the current iterate. Combining the extrapolation with the Newton step yields

$$w(0,0) \approx w(\rho,r) + (0-\rho)\dot{w}_\rho + (0-r)\dot{w}_r = w(\rho,r) - (\nabla\Phi(w))^{-1}r. \tag{12}$$

A recommended implementation of the predictor-corrector strategy is the following, given $w = w(\rho, r)$:

1. estimate $w(0,0)$ using the combined formula (12);
2. estimate a target value ρ^+;
3. estimate $w(\rho^+, 0)$ using equation (10) and update w.

1.5 Specialized predictor-corrector for linear optimization

Now, let us provide some specific details taking into account the particular system (7). Then, w is a triplet (x, y, s). The predictor step in (12), $\Delta_p w = -(\nabla\Phi(w))^{-1}r$ is best split into $(\Delta_p x, \Delta_p y, \Delta_p s)^t = \Delta_p w$. Similarly, the corrector step in equation (10), $\Delta_c w = -(\nabla\Phi(w))^{-1}(r + \rho v)$ is split into $(\Delta_c x, \Delta_c y, \Delta_c s)^t = \Delta_c w$.

1. The bounds $x \geqslant 0$ and $s \geqslant 0$ are implicit in all this context. Therefore, the computation of the estimate of $w(0,0)$ may be improved by enforcing the bounds:

$$\alpha_x = \max\{\alpha : x + \alpha\Delta_p x \geqslant 0\} \tag{13}$$
$$\alpha_s = \max\{\alpha : s + \alpha\Delta_p s \geqslant 0\} \tag{14}$$

   Therefore, the predicted values are $x_p = x + \alpha_x\Delta_p x$ and $s_p = s + \alpha_s\Delta_p s$.

2. A recommended value for ρ^+ is $\sigma\frac{s^t x}{n}$ with $\sigma = \left(\frac{x_p^t s_p}{x^t s}\right)^3$.

3. Again the bounds $x \geqslant 0$ and $s \geqslant 0$ will be enforced (still denoted by α_x and α_s); a further safeguard is to take $\hat{\alpha}_s = \min(0.99\alpha_s, 1)$ and $\hat{\alpha}_x = \min(0.99\alpha_x, 1)$ to ensure interior points. The next iterate is

$$(s^+, y^+) = (s, y) + \hat{\alpha}_s(\Delta_c s, \Delta_c y) \tag{15}$$
$$x^+ = x + \hat{\alpha}_x\Delta_c x \tag{16}$$

1.5.1 Variants

Whenever the unit stepsizes are used for the predictor step, at the predicted point, the primal and dual residuals will vanish. Indeed, Newton's method solves the linearized problem, and

the two equations associated with those residuals are already linear. Therefore, the right hand side for the correction step would be

$$\begin{pmatrix} Ax_p - b \\ y_p A + s_p - c \\ S_p X_p e - \rho^+ \mathbb{1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ S_p X_p e - \rho^+ \mathbb{1} \end{pmatrix}. \tag{17}$$

When the unit stepsizes are not used, one may keep the above right hand side or put in the (presumably very small) predicted primal and dual residuals.

1.5.2 Linear algebra issues

The tremendous success of predictor-corrector approaches is closely related to very clever treatment of the matrices underlying the linear systems. Observe that once a factorization of the matrix

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^t & I \\ S & 0 & X \end{bmatrix}$$

is completed, it will be reused twice, once to compute the prediction and once more to compute the correction.