The Penalty Interior Point Method fails to converge for Mathematical Programs with Equilibrium Constraints*

Sven Leyffer†

4 February 2002

Abstract

This paper presents a small example for which the Penalty Interior Point Method converges to a non-stationary point. The reasons for this adverse behaviour are discussed.

Keywords: Nonlinear programming, Interior Point Methods, PIPA, MPEC, MPCC, equilibrium constraints.

AMS-MSC2000: 90C30, 90C33, 90C51, 49M37, 65K10.

1 Introduction

The Penalty Interior Point Algorithm (PIPA) of Luo, Pang and Ralph [2, Chapter 6.1] solves Mathematical Programs with Equilibrium Constraints (MPECs) of the form

$$\begin{array}{ll} \text{minimize} & f(x,y,w,z) \\ \text{subject to} & x \in X \\ & F(x,y,w,z) = 0 \\ & 0 \le y \ \perp \ w \ge 0, \end{array} \tag{1.1}$$

where $X \subset \mathbb{R}^n$ is a polyhedral set, f and F are twice continuously differentiable functions and $y, w \in \mathbb{R}^m$.

In the remainder, d will denote the step or displacement computed by PIPA and subscripts like d_x will refer the part of d corresponding to the x-variables. Superscripts are used to denote iterates or evaluation of functions at a particular point, e.g. $\nabla f^k = \nabla f(x^k, y^k, w^k, z^k)$. Diagonal matrices are denoted by $W = \text{diag}(w)$ and $Y = \text{diag}(y)$ and $e = (1, \ldots, 1)^T$.

The algorithm solves the following direction finding problem at every iteration

$$\begin{array}{ll} \text{minimize} & \nabla f^{k^T} d + \frac{1}{2} d_x^T W^k d_x \\ \text{subject to} & x^k + d_x \in X \\ & F^k + \nabla F^{k^T} d = 0 \\ & Y^k d_w + W^k d_y = -Y^k w^k + \sigma \frac{y^{k^T} w^k}{m} e \\ & \|d_x\|^2 \le c \left(\|F^k\| + y^{k^T} w^k \right), \end{array} \tag{1.2}$$

*Numerical Analysis Report NA/208, Department of Mathematics, University of Dundee.

†Department of Mathematics, University of Dundee, (sleyffer@maths.dundee.ac.uk). Supported by EPSRC grant GR/M59549.

where $c > 0$ and $\sigma \in (0,1)$ are parameters. PIPA can be summarized as follows.

Penalty Interior Point Algorithm (PIPA)

1. *Initialization:*
   Choose parameters $c > 0$, $\sigma, \gamma, \rho \in (0,1)$,
   choose a starting point (x^0, y^0, w^0, z^0) such that $y^0, w^0 > 0$ suitably centered,
   choose a penalty parameter $\alpha > 0$ and set $k = 0$.

REPEAT

2. *Direction finding problem:* Solve problem (1.2) for a trial step $d = d^k$.
3. *Step size determination:* Find a step size $\tau = \tau_k$ to:
   3.1 Ensure centrality and positivity of (y, w) by finding the root of $g_k(\tau)$ in (1.3) or setting $\tau = 1 - \epsilon$ if this root does not exist or is greater than 1.
   3.2 Ensure sufficient reduction in the quadratic penalty function $P_\alpha(x, y, w, z)$ in (1.4) by performing an Armijo-search on P_α.
   Let τ_k be the step size determined in 3.1 and 3.2.
4. *Update:* $(x^{k+1}, y^{k+1}, w^{k+1}, z^{k+1}) = (x^k, y^k, w^k, z^k) + \tau_k(d_x^k, d_y^k, d_w^k, d_z^k)$, $k = k + 1$.

UNTIL $\|d\| \leq \epsilon$

The step size selection in Step 3. requires two functions, namely

$$g_k(\tau) = (1-\rho)\sigma \frac{{y^k}^T w^k}{m} + \tau \left(\min_{1 \leq i \leq m} d_{y_i}^k d_{w_i}^k - \rho \frac{{d_y^k}^T d_w^k}{m} \right) \qquad (1.3)$$

which controls the rate at which y, w are changed and the quadratic penalty function

$$P_\alpha(x, y, w, z) = f(x, y, w, z) + \alpha \left(\|F(x, y, w, z)\|^2 + y^T w \right). \qquad (1.4)$$

More details of the algorithm can be found in [2, Chapter 6.1]. Note, that this presentation of PIPA is less sophisticated than [2] which allows for instance σ to vary with the iteration.

Convergence of PIPA is established in [2] under the following two assumptions.

[SC] Strict complementarity of the solution $y^* + w^* > 0$.

[NS] Nonsingularity of the matrix

$$\begin{bmatrix} \nabla_y F^* & \nabla_w F^* \\ W^* & Y^* \end{bmatrix}.$$

Theorem 1.1 [2, Theorem 6.1.17] Suppose that the Hessian matrices W^k are bounded and that $\sigma > 0$. If the penalty parameter α is bounded, then every limit point of (x^k, y^k, w^k, z^k) that satisfies **[SC]** and **[NS]** is a stationary point of (1.1).

In the next section, a small example is presented for which PIPA converges to a non-stationary point, contradicting Theorem 1.1.

2 A cautionary example

This section examines the behaviour of PIPA applied to the following small example

$$\begin{array}{ll} \text{minimize} & x + w \\ \text{subject to} & -1 \leq x \leq 1 \\ & -1 + x + y = 0 \\ & 0 \leq y \ \perp \ w \geq 0. \end{array} \tag{2.1}$$

The solution of (2.1) is $(x, y, w)^* = (-1, 2, 0)$. This solution satisfies the assumptions **[SC]** $(y^* + w^* = 2 > 0)$ and **[NS]**, since

$$\begin{bmatrix} \nabla_y F^* & \nabla_w F^* \\ W^* & Y^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

is nonsingular, so that the assumptions of [2, Theorem 6.1.17] are satisfied.

2.1 Initialization

Let the starting point be $(x^0, y^0, w^0) = (0, 1, 0.02)$ and assume that the following parameters are chosen: $c = 1, \sigma = 0.1, \gamma = 0.01, \rho = 0.9$ and $\alpha = 2$. Note, that the starting point satisfies the linear equation, which implies that $\|F^k\| = 0$ for all subsequent iterations.

2.2 Direction Finding Problem

The direction finding problem for this example can be simplified as follows

$$\begin{array}{ll} \text{minimize} & d_x + d_w \\ \text{subject to} & -1 \leq x^k + d_x \leq 1 \\ & d_x + d_y = 0 \\ & w^k d_y + y^k d_w = (\sigma - 1) y^k w^k \\ & -\sqrt{y^k w^k} \leq d_x \leq \sqrt{y^k w^k} \end{array}$$

The aim is to show that the presence of the trust-region like bound on d_x can under certain circumstances prevent convergence to a stationary point.

Note that the direction finding problem is an LP. It can be seen that at the solution of this LP, the two equations and the lower bound $-\sqrt{y^k w^k} \leq d_x$ are active. Thus the optimal basis is given by

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & w^k & 0 \\ 0 & y^k & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 & -\frac{w^k}{y^k} \\ 0 & 0 & \frac{1}{y^k} \\ 1 & -1 & \frac{w^k}{y^k} \end{bmatrix}.$$

From this, multipliers can be computed as

$$\lambda^k = B^{-1} c = \begin{bmatrix} 0 & 1 & -\frac{w^k}{y^k} \\ 0 & 0 & \frac{1}{y^k} \\ 1 & -1 & \frac{w^k}{y^k} \end{bmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\frac{w^k}{y^k} \\ \frac{1}{y^k} \\ 1 + \frac{w^k}{y^k} \end{pmatrix}.$$

and optimality follows, since $\lambda_3^k \geq 0$ for the only inequality constraint. The step d can also be computed as

$$d^k = B^{-T} b^k = \begin{pmatrix} -\sqrt{y^k w^k} \\ \sqrt{y^k w^k} \\ -\frac{9}{10} w^k - \frac{w^k}{y^k}\sqrt{y^k w^k} \end{pmatrix},$$

where $b^k = (0, -\frac{9}{10} y^k w^k, -\sqrt{y^k w^k})^T$ is the right hand side of the active constraints.

2.3 Step size determination

The step size rule ensures that $\tau \leq 1$. For this example, the root of $g_k(\tau)$ in (1.3) can be shown to be given by

$$\tau = -\frac{1}{10}\frac{y^k w^k}{d_w^k d_y^k} = \frac{(y^k)^2}{\left(9y^k + 10\sqrt{y^k w^k}\right)\sqrt{y^k w^k}}.$$

Expanding τ in terms of powers of w around $w = 0$ (using `maple`'s `series` function) gives

$$\tau = \frac{1}{9}\frac{\sqrt{y}}{\sqrt{w}} - \frac{10}{81} + \frac{100}{729}\frac{\sqrt{w}}{\sqrt{y}} - \frac{1000}{6561}\frac{w}{y} + \mathcal{O}(w^{3/2}).$$

And it can be seen that, if $y \geq 1$ and $w \leq 0.02$, then τ can be bounded by

$$\tau \geq \frac{10\sqrt{2}}{18} - \frac{10}{81} - \frac{200}{6591} + o(w).$$

Evaluating the right hand side shows it to be roughly 0.6319, so that the following bounds on τ hold

$$\frac{5}{9} \leq \tau \leq 1. \tag{2.2}$$

Below it will be shown that this step size is also acceptable in the Armijo test.

2.4 Update of variables

Since $\tau \leq 1$, it follows that

$$x^{k+1} = x^k - \tau\sqrt{y^k w^k} \geq x^k - \sqrt{y^k w^k} \tag{2.3}$$

$$y^{k+1} = y^k + \tau\sqrt{y^k w^k} \leq y^k + \sqrt{y^k w^k}, \tag{2.4}$$

while $\tau \geq \frac{5}{9}$ implies that

$$w^{k+1} = w^k + \tau\left(-\frac{9}{10}w^k - \frac{w^k}{y^k}\sqrt{y^k w^k}\right) \leq \frac{1}{2}w^k - \frac{5}{9}\frac{w^k}{y^k}\sqrt{y^k w^k}. \tag{2.5}$$

Thus it is possible to bound the new complementarity violation after the step by

$$\begin{aligned} y^{k+1}w^{k+1} &\leq \left(y^k + \sqrt{y^k w^k}\right)\left(\tfrac{1}{2}w^k - \tfrac{5}{9}\tfrac{w^k}{y^k}\sqrt{y^k w^k}\right) \\ &= \tfrac{1}{2}y^k w^k - \tfrac{5}{9}w^k\sqrt{y^k w^k} + \tfrac{1}{2}w^k\sqrt{y^k w^k} - \tfrac{5}{9}(w^k)^2 \\ &\leq \tfrac{1}{2}y^k w^k. \end{aligned}$$

It also follows, that $y^{k+1} \geq y^k \geq 1$ and that $w^{k+1} \leq \frac{1}{2}w^k \leq 0.02$ so that the conditions for (2.2) remain satisfied.

2.5 Penalty reduction

The penalty update rule in [2, (6.1.24)] finds the smallest integer $p \geq 1$ such that

$$\nabla f^{k^T} d - \alpha^p (1-\sigma) y^{k^T} w^k \;<\; -\alpha^p (1-\sigma) y^{k^T} w^k \;<\; -y^{k^T} w^k$$

holds. Since $\nabla f^{k^T} d < 0$ and $\alpha(1-\sigma) > 1$, this is always satisfied for $p = 1$ and the penalty parameter is never increased.

It remains to show that the step size (2.2) satisfies the conditions for the Armijo search in Step 3.2. It can be seen that the actual reduction satisfies

$$P_\alpha^{k+1} - P_\alpha^k = \tau d_x + \tau d_w + \alpha \left(y^{k+1} w^{k+1} - y^k w^k \right).$$

Since $y^{k+1} w^{k+1} \leq \frac{1}{2} y^k w^k$ it follows that

$$P_\alpha^{k+1} - P_\alpha^k \leq \tau d_x + \tau d_w - \alpha \frac{1}{2} y^k w^k,$$

which implies the sufficient reduction condition,

$$P_\alpha^{k+1} - P_\alpha^k \leq \gamma \tau d_x + \gamma \tau d_w - \alpha(1-\sigma)\gamma y^k w^k,$$

since $\gamma \leq 1$ and $(1-\sigma)\gamma \leq \frac{1}{2}$. Thus, each step of PIPA generates new iterates which satisfy the bounds (2.3), (2.4) and (2.5).

2.6 Limit and numerical experiment

A lower bound on the global changes in the controls x can now be given

$$x^{k+1} \;=\; x^k - \tau_k \sqrt{y^k w^k} \;=\; x^0 - \sum_{l=0}^{k} \tau_l \sqrt{y^l w^l} \;\geq\; x^0 - \sum_{l=0}^{k} \sqrt{y^l w^l}$$

where the last inequality follows from $\tau_l \leq 1$. Next, using the fact that $x^0 = 0$ and the bound $y^l w^l \leq \left(\frac{1}{2}\right)^l y^0 w^0$ and $y^0 w^0 = \frac{2}{100}$, it follows that

$$x^{k+1} \;\geq\; x^0 - \sum_{l=0}^{k} \left(\frac{1}{\sqrt{2}} \right)^l \sqrt{y^0 w^0} \;=\; \frac{-\sqrt{2}}{10} \left(\frac{1 - \left(\frac{1}{\sqrt{2}}\right)^k}{1 - \frac{1}{\sqrt{2}}} \right).$$

By taking the limit $(k \to \infty)$ on the right, it follows, that all control iterates satisfy the bound

$$x^{k+1} \;\geq\; \frac{-2}{10\left(\sqrt{2} - 1\right)} \approx -0.4828. \tag{2.6}$$

Similarly, an upper bound can be derived for the y-iterates as

$$y^{k+1} \;\leq\; y^0 + \sum_{l=0}^{k} \left(\frac{1}{\sqrt{2}} \right)^l \sqrt{y^0 w^0} \;=\; 1 + \frac{\sqrt{2}}{10} \left(\frac{1 - \left(\frac{1}{\sqrt{2}}\right)^k}{1 - \frac{1}{\sqrt{2}}} \right),$$

k	x^k	y^k	w^k	ared	pred
1	0	1	0.02		
2	-0.096022613	1.0960226	0.0058578644	-0.198	-0.137
3	-0.17606958	1.1760696	0.00016323495	-0.0974	-0.0982
4	-0.18991126	1.1899113	1.4549224E-05	-0.0143	-0.0143
5	-0.1940679	1.1940679	1.4171928E-06	-0.00421	-0.0042
6	-0.19536745	1.1953675	1.4145236E-07	-0.00131	-0.0013
7	-0.19577825	1.1957782	1.4223933E-08	-0.000412	-0.000411
8	-0.19590853	1.1959085	1.433645E-09	-0.00013	-0.00013
9	-0.1959499	1.1959499	1.4460519E-10	-4.14e-05	-4.14e-05
10	-0.19596304	1.195963	1.4586827E-11	-1.32e-05	-1.31e-05

Table 1: Iterates, actual and predicted reduction of AMPL implementation of PIPA

which shows that

$$y^{k+1} \le 1 + \frac{2}{10\left(\sqrt{2}-1\right)} \approx 1.4828. \qquad (2.7)$$

These bounds can also be verified numerically.

Example (2.1) and PIPA has been implemented in AMPL [1] using the "looping extension" which allows the convenient implementation of algorithms in AMPL. The sequence that is generated is given in Table 1 and confirms the bounds given above.

2.7 A counter example for PIPA

From the previous sections it follows that all iterates remain in a compact set, namely

$$\begin{array}{rcccccl} 0 & = & x^0 & \ge & x^k & \ge & -0.4828 \\ 1 & = & y^0 & \le & y^k & \le & 1.4828 \\ \frac{2}{100} & = & w^0 & \ge & w^k & \ge & 0. \end{array}$$

Thus, the sequence $(x,y,w)^k$ has a limit point $(x,y,w)^\infty$. From (2.5) it follows that

$$w^{k+1} \le w^k \left(\frac{1}{2} - \frac{5}{9y^k}\sqrt{y^k w^k} \right) \le \frac{1}{2} w^k$$

so that $w^\infty = 0$ and ${y^\infty}^T w^\infty = 0$. Therefore, PIPA converges to a limit point $(x,y,w)^\infty$ such that

$$\begin{array}{rcccl} 0 & \ge & x^\infty & \ge & -0.4828 \\ 1 & \le & y^\infty & \le & 1.4828 \\ & & w^\infty & = & 0. \end{array}$$

This limit point satisfies complementarity and the linear constraints for problem (2.1) (since $(x,y,w)^0$ satisfies the linear constraints). Moreover, this limit point also satisfies the assumptions of Theorem 1.1. Assumption **[SC]** is satisfied, since $y^\infty + w^\infty \ge 1 > 0$ and Assumption **[NS]** holds, since

$$\begin{bmatrix} \nabla_y F^\infty & \nabla_w F^\infty \\ W^\infty & Y^\infty \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & y^\infty \end{bmatrix}$$

is nonsingular for all $1 \le y^\infty \le 1.4828$. However, the limit point is clearly not a stationary point, thereby contradicting Theorem 1.1.

3 Conclusion and Discussion

In the previous section, a small example has been presented for which PIPA converges to feasible but non-stationary limit point. This example seems to contradict [2, Theorem 6.1.17] which establishes the convergence of PIPA.

The reason for this apparent failure of PIPA is the "trust-region" type constraint

$$\|d_x\|^2 \;\leq\; c\left(\|F^k\| + y^{k^T} w^k\right) \;=:\; \Delta_k \tag{3.8}$$

in the direction finding problem (1.2). The trust-region radius Δ_k converges to zero as the iterates approach feasibility, thereby limiting the progress towards optimality in the controls x. This adverse behaviour can be expected to occur whenever the iterates approach feasibility "faster" than optimality.

Linking the trust-region radius to the feasibility of the problem is counter intuitive. Normally, Δ_k is controlled by the algorithm, rather than by the iterates. In particular, it should reflect how well the model problem (1.2) approximates the original problem, measured for instance by the agreement between actual and predicted reduction. In this example, actual and predicted reduction show near perfect agreement towards the end (see last two columns in Table 1). Thus one would expect Δ_k to be increased rather than decreased.

The adverse situation in which (3.8) limits progress towards optimality can be easily detected. The Lagrange multiplier of (3.8) indicates whether or not progress can be made by relaxing this constraint. In the present example, this multiplier converges to 1.

Acknowledgements

I am grateful to Roger Fletcher and Jorge Nocedal for many fruitful discussions on MPECs and interior point methods. I am also grateful to Jong-Shi Pang for going over an earlier version of this paper with me.

References

[1] Fourer, R., Gay, D.M. and Kernighan, B.W. *AMPL: A modelling Language for Mathematical Programming.* boyd & fraser publishing company, Massachusetts, 1993.

[2] Luo, Z.-Q., Pang, J.-S. and Ralph, D. *Mathematical Programs with Equilibrium Constraints.* Cambridge University Press, 1996.