

A one phase IPM for non-convex optimization

Oliver Hinder, Yinyu Ye

June 24, 2017

Abstract

Based on the work of [REF] it has been assumed that infeasible start IPM developed for conic optimization [REF] cannot be adapted to non-linear optimization and converge without significant modification i.e. a two phase method or a penalty function. We show that, in fact, by careful initialization and non-linear updates on the primal variables.

1 Introduction

[?] showed that if we apply an infeasible start IPM to:

$$\begin{aligned} \min x \\ x^2 - s_1 - 1 &= 0 \\ x - s_2 - 1/2 &= 0 \\ s_1, s_2 &\geq 0 \end{aligned}$$

Fails to converge to either a local optimum or infeasibility certificate

Modifications:

- A. Two phases (IPOPT)
- B. Compute two directions (KNITRO)
- C. Penalty (or big-M) method e.g. [?, ?]

1.1 Relevant literature

Within non-convex optimization there are four papers that I think are particularly relevant to our work:

- A. The paper [?] shows that there are examples for which infeasible start algorithms will always fail to converge to either a optimal solution or a stationary measure of infeasibility when constraints are non-convex (irrespective of the strategy for used). This is the inspiration for the two phase algorithm of IPOPT and justifies why our one phase algorithm is necessary.
- B. The description of the IPOPT algorithm [?]. IPOPT uses a two phase method the primary phase searches simultaneously for optimality and feasibility using a classical infeasible start method and a feasibility restoration phase that minimizes infeasibility. The feasibility restoration phase is only called when the step size for the infeasible start method is small. Another distinct feature of the algorithm is the filter line search (which allows progress on either the constraints or the objective).
- C. The description of the KNITRO algorithm [?]. KNITRO is a trust region algorithm. The approach is quite distinct from typical infeasible start algorithms and is worth looking at (each step computes two different directions, using two different linear systems, one to reduce the objective and the other to reduce infeasibility). There is a more recent paper [?] that adds an feasibility restoration phase (this is theoretically unnecessary, but the practical results are good).
- D. The paper [?] introduces a barrier penalty method. This paper uses a similar approach to us. The main different with our approach is we treat λ as a dual variable, whereas in Curtis's paper λ is replaced by a penalty parameter that is updated in an ad hoc fashion.
- E. Papers in convex optimization?

2 Motivation and derivation of algorithm

We wish to solve the following problem:

$$\begin{aligned} \min f(x) \\ a(x) \leq 0 \end{aligned}$$

Assume:

- A. The constraints and objective are C^2 - L_0, L_1 etc
- B. The set $a(x) \leq \theta$ is bounded for any $\theta \in \mathbb{R}^m$

2.1 Infeasibility certificates

First-order local L1-infeasibility certificate, if x^* is a first-order local optimum of:

$$\begin{aligned} \min e^T z \\ a(x) - s + z = 0 \\ s, z \geq 0 \end{aligned}$$

With $e^T z > 0$.

First-order local farkas infeasibility certificate, if there exists some $w \leq 0$, such that x^* is a first-order local optimum of:

$$\min w^T a(x)$$

With $w^T a(x) > 0$. Now, suppose we have found a point such that

$$\|w^T \nabla a(x)\|_2 < \frac{w^T a(x)}{\|x - x^*\|_2}$$

Now, if the function $w^T a(x)$ is convex then we have:

$$w^T a(x^*) \geq w^T a(x) + w^T \nabla a(x)(x - x^*) > 0$$

Therefore we declare a problem locally infeasible if our algorithm finds a point with:

$$\frac{\|w^T \nabla a(x)\|_2}{w^T a(x)} \leq \epsilon_{\text{inf}}$$

Where ϵ_{inf} is the tolerance for primal feasibility.

$L_\infty \dots$

2.2 Derivation of method

Log barrier problem:

$$\begin{aligned} \min f(x) - \mu^k \sum_i \log(s_i) \\ a(x) + s = 0 \\ s \geq 0 \end{aligned}$$

Shifted log barrier problem

$$\min f(x) - (1 - \eta^k) \mu^k \sum_i \log s_i + \frac{\delta^k}{2} \|x - x^k\|^2 \quad (1)$$

$$a(x) + s = (1 - \eta^k)(a(x^k) + s^k) \quad (2)$$

$$s \geq 0 \quad (3)$$

KKT system:

$$\begin{aligned}\nabla f(x) + \delta^k(x - x^k) + y^T \nabla a(x) &= 0 \\ a(x) + s &= (1 - \eta^k)(a(x^k) + s^k) \\ s_i y_i &= (1 - \eta^k) \mu^k \\ s, y &\geq 0\end{aligned}$$

Corresponding newton system:

$$\begin{bmatrix} \nabla_x^2 L(x^k, y^k) + \delta^k I & \nabla a(x^k)^T & 0 \\ \nabla a(x^k) & 0 & I \\ 0 & S^k & Y^k \end{bmatrix} \begin{bmatrix} d_x^k \\ d_y^k \\ d_s^k \end{bmatrix} = \begin{bmatrix} -(\nabla f(x^k) + \nabla a(x^k)^T y^k) \\ -\eta^k(a(x^k) + s^k) \\ (1 - \eta^k) \mu^k e - Y^k s^k \end{bmatrix} \quad (4)$$

By taking the primal schur complement one can see solving system (??) is equivalent to solving:

$$(M(x^k, y^k, s^k) + \delta I) d_x^k = -\nabla f(x^k) - \nabla a(x^k)^T ((1 - \eta^k) \mu^k (S^k)^{-1} e + \eta^k Y^k (e + (S^k)^{-1} (a(x^k) + s^k)))$$

Where:

$$M(x, y, s) = \nabla^2 \mathcal{L}(x, y) + \nabla a(x)^T Y S^{-1} \nabla a(x) \quad (5)$$

$$D(x, y) \leq \mu^k \& \|Sy - \mu^k\|_\infty \leq \mu^k / 4 \quad (6)$$

2.3 Behavior of this algorithm on convex quadratic programs

[one stabilization step followed by one aggressive step]. [resemblance to predictor corrector]

3 Practical algorithm

$$\sigma(y) = \frac{100}{\max\{100, \|y\|_\infty\}}$$

Scaled KKT termination criterion:

$$\sigma(y) \|\nabla L_\gamma(x, y)\|_\infty \leq \epsilon \quad (7a)$$

$$\sigma(y) \|Sy\|_\infty \leq \epsilon \quad (7b)$$

$$\|a(x) + s\|_\infty \leq \epsilon \quad (7c)$$

Local primal infeasibility termination criterion:

$$\frac{\|w^T \nabla a(x)\|_2}{w^T a(x)} \leq \epsilon_{\text{inf}} \quad (8a)$$

$$\frac{\|w^T \nabla a(x)\|_\infty}{\|w\|_\infty} \leq \epsilon_{\text{inf}} \quad (8b)$$

Local dual infeasibility termination criterion

$$\frac{\|w^T \nabla a(x)\|_2}{w^T a(x)} \leq \epsilon_{\text{inf}} \quad (9a)$$

$$\frac{\|w^T \nabla a(x)\|_\infty}{\|w\|_\infty} \leq \epsilon_{\text{inf}} \quad (9b)$$

Aggressive step criterion:

$$\|\nabla L_\gamma(x, y)\|_\infty \leq \min \left\{ \frac{\|a(x) + s\|_\infty}{\sigma(y)}, 1 + \|\nabla f_\gamma(x)\|_\infty \right\} \quad (10a)$$

$$\frac{Sy}{\mu} \in [e\beta_2, e/\beta_2] \quad (10b)$$

Algorithm 1 High level description of one phase IPM

Input: Some initial point x_0, y_0, s_0, μ

For each iteration $i = 1, \dots, i_{\max}$ perform the following steps:

A.1 Update hessian and jacobian.

A.2 Check the termination criterion given in (??), (??) and (??).

A.3 Form primal schur complement at the point x, y

A.4 Factorize the matrix using IPOPT strategy

A.5 Perform corrections steps for $k \in \{1, \dots, c_{\max}\}$

A.5.1 Perform corrections

-Case-I If equation (??) is satisfied then go to the aggressive search direction algorithm (Algorithm ??).

-Case-II Otherwise go to the stablization search direction algorithm (Algorithm ??).

A.5.2 Deal with failures. If correction succeeds set $x, y, s \leftarrow x^+, y^+, s^+$. If correction failed and $k = 1$ go to (??). Otherwise if the correction failed with $k > 1$ go to step (??).

A.6 Go to (??)

A.7 Increase delta to address failure. Set $\delta = \max\{10\delta, \delta_{\min}\}$ and factorize the matrix. Go to step (??).

3.1 Updating the iterates

Given a primal step size α_P we update the primal iterates as follows:

$$x^+ \leftarrow x + \alpha_P d_x \quad (11a)$$

$$\mu^+ \leftarrow (1 - \eta\alpha_P)\mu \quad (11b)$$

$$\theta^+ \leftarrow (1 - \eta\alpha_P)\theta \quad (11c)$$

$$s^+ \leftarrow a(x^+) + \theta(s^1 - a(x^1)) \quad (11d)$$

Fraction to the boundary:

$$s + \alpha_P d_s \geq s\beta_3 \quad (12)$$

Given some candidate primal iterate x^+, s^+ , we define $B(s^+, d_y)$ set of feasible dual step sizes. In particular, let it be the largest interval such that if $\alpha_D \in B(s^+, d_y)$ then

$$\frac{S^+ y + \alpha_D d_y}{\mu} \in [\beta_1, 1/\beta_1] \quad (13)$$

If no such interval exists we set $B(s^+, d_y)$ to the empty set and the step will be rejected. We compute the dual step size as follows:

$$\alpha_D = \arg \min_{\alpha_D \in B(s^+, d_y)} \|S^+ y + \alpha_D S^+ d_y\|_2^2 + \sigma(y) \|\nabla L(x, y) + (\nabla^2 L(x, y) + \delta I) d_x + \alpha_D \nabla a(x)^T d_y\|_2^2 \quad (14)$$

This reduces to a one dimensional least squares problem in α_D which has a closed form expression.

3.2 Aggressive search directions

Algorithm 2 High level description of aggressive correction

Input: Some point x, y, s, μ

Output: A new point x^+, y^+, s^+, μ^+ and a **status**

A.1 Compute a direction d from the system (??) with $\eta = 1$

A.2 Estimate the largest primal step size α_P^{\max}

A.3 *Perform a backtracking line search on the primal step α_P .* Trial step sizes $\alpha \in \{\alpha_P^{\max}, \beta\alpha_P^{\max}, \dots\}$ and terminate with **status** = SUCCESS the first time all of the following conditions hold:

- (i) The fraction to the boundary rule (??) is satisfied
- (ii) The set of valid dual step sizes is non-empty i.e. $B(s^+, d_y) \neq \emptyset$
- (iii) Lagrangian???

Terminate with **status** = FAILURE if the line search trials a point with

$$\alpha_P \leq \text{some constant} \times \min_{\{j: a_j(x) + s_j < 0\}} \frac{-s_j}{a_j(x) + s_j}$$

3.3 Stabilization search directions

Let:

$$\mathbb{K}(x, y, s) = \sigma(y) \max\{\|\nabla L(x, y)\|_\infty, \|Sy - \mu\|_\infty\} \quad (15)$$

During the stabilization search we accept iterates that sufficiently reduce the merit function:

$$\phi(x^+, y^+, s^+) \leq \phi(x, y, s) + \beta_6 \tilde{\Delta}_{(x, y, s)}^\phi(d_x, d_y) \quad (16)$$

Or sufficient progress is made on a KKT merit function relative to the historic iterates. In particular, for every $(\hat{x}, \hat{y}, \hat{s}) \in \mathbb{H}$ one of the following two equations holds:

$$\mathbb{K}(x^+, y^+, s^+) \leq (1 - \beta_5 \alpha_P) \mathbb{K}(\hat{x}, \hat{y}, \hat{s}) \quad (17a)$$

$$\phi(x^+, y^+, s^+) \leq \phi(\hat{x}, \hat{y}, \hat{s}) - \beta_5 \alpha_P \mathbb{K}(\hat{x}, \hat{y}, \hat{s}). \quad (17b)$$

This filter accepts large steps that significantly reduce the KKT error. The motivation for this choice is that, even on non-convex problems, the optimization trajectory often passes through regions where the primal schur complement is positive definite. For example, this naturally occurs near points satisfying the sufficient conditions for local optimality. In these situation often the KKT error is a better measure of progress towards a local optimum than a merit function that discards information about the dual feasibility.

Algorithm 3 High level description of stabilization corrections

Input: Some point x, y, s, μ

A.1 Compute a direction d from the system (??) with $\eta = 0$

A.2 Estimate the largest primal step size α_P^{\max}

A.3 *Perform a backtracking line search on the primal step α_P .* Start with $\alpha_P = \alpha_P^{\max}$. Set **status** = SUCCESS and move to step (??) if all of the following conditions hold:

- (i) The fraction to the boundary rule (??) is satisfied
- (ii) The set of valid dual step sizes is non-empty i.e. $B(s^+, d_y) \neq \emptyset$
- (iii) The filter i.e. one of equations (??) is satisfied.

Return the original iterates and **status** = FAILURE if the line search trials a point with $\alpha_P \leq \beta_4$

A.4 Augment the histories \mathbb{H} with the new iterate (x^+, y^+, s^+) .

Table 1 Parameters values and descriptions

Parameter	Description	Possible values	Chosen value
c_{\max}	Maximum number of corrections per iteration. See (??).	Any natural number	3
β_1	Restricts how far complementarity of s and y can be from μ . See (??).	The interval $(0, 1)$	0.01
β_2	Restricts how far complementarity of s and y can be from μ in order for the aggressive criterion to be met. See (??).	The interval $(0, \beta_1)$	
β_4	Minimum step size for stable line searches. See (??).	The interval $(0, 1)$	
β_5	Acceptable reduction factor for the scaled KKT error \mathbb{K} during stabilization steps. See (??).	The interval $(0, 1)$	
β_6	Acceptable reduction factor for the merit function ϕ during stabilization steps. See (??).	The interval $(0, 1)$	

4 Convergence proofs

Lemma 1. Assume that $a(x^1) - s^1 = -\mu e$ and $\epsilon < 1$. If the criterion for an aggressive step is met at iteration k then we have:

$$\|y^k\|_1 \geq \frac{\|\nabla c(x^k)\|_2}{\epsilon^2} + 3m$$

Proof. Observe that:

$$-a(x)^T y = -(a(x) - s)^T y - s^T y \geq \mu(e^T y - 2)$$

Therefore:

$$\frac{\|\nabla a(x)^T y\|}{-a(x)^T y} \leq \frac{\mu^k \sqrt{\|y\|_1 + 1} + \|\nabla c(x)\|}{\mu(\|y\|_1 - 2m)}$$

If:

$$\|y^k\|_1 \geq \frac{\|\nabla c(x^k)\|_2 + 3m}{\epsilon^2}$$

Then:

$$\frac{\|\nabla a(x)^T y\|}{-a(x)^T y} \leq \epsilon$$

Which gives the result. \square

Proof. Observe that:

$$-a(x)^T y = -(a(x) - s)^T y - s^T y \geq \mu(e^T y - 2)$$

Therefore:

$$\frac{\|\nabla a(x)^T y\|}{-a(x)^T y} \leq \frac{1 + \|\nabla c(x)\|}{\mu(\|y\|_1 - 2m)}$$

If:

$$\|y^k\|_1 \geq \frac{\|\nabla c(x^k)\|_2}{\epsilon^2} + 3m$$

Then:

$$\frac{\|\nabla a(x)^T y\|}{-a(x)^T y} \leq \epsilon$$

Which gives the result. \square

Lemma 2. Assume that $a(x^1) - s^1 = -\mu e$ and $\epsilon < 1$. The algorithm, SIMPLIFIED-ONE-PHASE-NON-CONVEX-IPM, with any takes at most $\frac{\mu^0(2\|\nabla c(x^k)\|_2+8)}{\epsilon^2}$ aggressive steps to terminate with either a first-order stationary point or first-order farkas certificate.

Proof. We wish to prove that for any δ with

$$\delta \geq \frac{\|g^k\|_{L_0}}{\mu^k} - \lambda_{\min}(M^k)$$

and α satisfying

$$\alpha \leq \frac{1}{\|y^k\|_{\infty} + 4} \quad (18)$$

the iterate $x^+ = x^k + \alpha d_x^k$, $y^+ = y^k + \alpha d_y^k$, $\mu^+ = \mu(1 - \alpha)$ is feasible. Observe that this implies the result since if: $\alpha \geq \frac{1}{2(\|y^k\|_{\infty} + 4)}$ then:

$$\mu^{k+1} = (1 - \alpha)\mu^k = \mu^k - \frac{\mu^k}{2\|y^k\|_{\infty} + 8} \leq \mu^k - \frac{\epsilon^2}{2\|\nabla c(x^k)\|_2 + 8}.$$

We wish to show that $s^{k+1} \in [s^k/2, 3s^k/2]$. Where $s^{k+1} = a(x + \alpha_P d_x) + (1 - \alpha_P)\mu^k e$. Subtracting and adding $s^k = a(x^k) + \mu^k e$ yields

$$s^{k+1} = s^k + (a(x^k + \alpha_P^k d_x^k) - a(x^k)) - \alpha_P^k \mu^k e$$

Therefore, it remains to bound the term $a(x^k + \alpha_P^k d_x^k) - a(x^k) - \alpha_P^k \mu^k e$. Applying our assumption on α^k , we immediately get $0 \leq \alpha_P^k \mu^k e \leq s^k/4$. Furthermore, we know that $\|d_x^k\|_2 \leq \mu^k L_0$ therefore:

$$\alpha_P^k \|d_x^k\|_2 \leq \frac{\min_i \{s_i^k\}}{2L_0}$$

Since $a(x)$ is L_0 -Lipshitz we have:

$$-s^k/4 \leq a(x^k) - a(x^k + \alpha_P^k d_x^k) \leq s^k/4$$

which shows that $s^{k+1} \in [s^k/2, 2s^k]$. Observe that $y^{k+1} = y^k + \alpha^k d_y^k \geq y^k/2$. It remains to show that $\|y^{k+1} s^{k+1} - \mu^{k+1}\|_{\infty} \leq \mu^k/2$. Now we have:

$$d_y = -Y(S^{-1}d_s + e)$$

Hence using that $\|d_s\| \leq \dots$ we get $d_y \in [-2y, 2y]$. It follows that $y^k + \alpha_P d_y \in [y^k/2, 3y^k/2]$.

Finally, using the fact that $s^{k+1} \in s^k[3/4, 5/4]$ and $s^{k+1} \in y^k[3/4, 5/4]$ we have:

$$\frac{s^{k+1}y^{k+1}}{s^k y^k} \in [1/2, 3/2]$$

And since $\frac{s^k y^k}{\mu^k} \in [1/2, 3/2]$ we have $\frac{s^{k+1}y^{k+1}}{\mu^k} \in [1/4, 3]$ which concludes the proof. \square

4.1 Trust region

Algorithm 4 Stable-trust-region-step

```

function STABLE-TRUST-REGION-STEP( $f, a, x, y, r$ )
  for  $j = 1, \dots, \infty$  do
    ( $d_x, d_y, M^+$ )  $\leftarrow$  APPROX-PRIMAL-DUAL-TRUST-REGION( $f, a, x, y, r$ )
     $x^+, y^+, \alpha^+ \leftarrow$  stable-line-search(...)
    if  $\alpha^+ > \alpha_{\min}$  then
      if  $j = 1$  &  $\alpha = 1$  then
         $r^+ \leftarrow 10r$ 
      else if  $\alpha^+ < \alpha_{small}$  then
         $r^+ \leftarrow r/2$ 
      else
         $r^+ \leftarrow \|d_x\|_2$ 
      end if
      break
    end if
     $r^+ \leftarrow r^+/8$ 
  end for
  return ( $x^+, y^+, M^+, r^+$ )
end function

```

Algorithm 5 Stable-trust-region-step

```

function STABLE-TRUST-REGION-IPOPT-STYLE( $f, a, x, y, r$ )
  for  $j = 1, \dots, \infty$  do
    Factorize  $M$  with ...  $\delta$  ...
    if inertia is good then
      ...
    end if
     $x^+, y^+, \alpha^+ \leftarrow \text{stable-line-search}(\dots)$ 
  end for
  return  $(x^+, y^+, M^+, r^+)$ 
end function

```

4.2 Line searches

```

function MOVE( $f, a, x, y, d_x, d_y, \eta, \alpha_P$ )
   $x^+ \leftarrow x + \alpha_P d_x$ 
   $\mu^+ \leftarrow (1 - \eta \alpha_P) \mu$ 
   $\theta^+ \leftarrow (1 - \eta \alpha_P) \theta$ 
   $s^+ \leftarrow a(x^+) + \theta(s^1 - a(x^1))$ 
   $\alpha_D \leftarrow \arg \max_{\alpha \in [0,1]} \alpha$  s.t.  $\frac{S^+(y+d_y \alpha_D)}{\mu} \in [e\beta_1, e/\beta_1]$ 
   $y^+ \leftarrow y + \alpha_D d_y$ 
end function

```

```

function DUAL-LINE-SEARCH( $f, a, x^+, s^+, y, d_y, \alpha_P$ )
   $u_D, \text{status1} \leftarrow \arg \max_{\alpha \in [0,1]} \alpha$  s.t.  $\frac{S^+(y+d_y \alpha_D)}{\mu} \in [e\beta_1, e/\beta_1]$ 
   $l_D, \text{status2} \leftarrow \arg \min_{\alpha \in [0,1]} \alpha$  s.t.  $\frac{S^+(y+d_y \alpha_D)}{\mu} \in [e\beta_1, e/\beta_1]$ 
   $\alpha_D \leftarrow \arg \min_{\alpha \in [l_D, u_D]} D_\gamma(x^+, y + \alpha d_y)$ 
   $y^+ \leftarrow y + \alpha_D d_y$ 
end function

```

Algorithm 6 Stable line search

```

function STABLE-LINE-SEARCH( $f, a, x, y, s, d_x, d_y, d_s \eta$ )
   $\eta \leftarrow 0$ 
   $\alpha_P \leftarrow \text{FRACTIONTOBOUNDARY}(y, s, d_y, d_s)$ 
  for  $i = 1, \dots, \infty$  do
     $x^+, y^+, \text{status} \leftarrow \text{MOVE}(f, a, x, y, d_x, d_y, \eta, \alpha_P)$ 
    if status = feasible then
      if sufficient progress on merit function then
        return  $x^+, y^+$ 
      else
        end if
      else
         $\alpha_P \leftarrow \alpha_P / 2$ 
      end if
    end for
  end function

```

5 Scrap paper

6 Log barrier sub-problems

This paper is concerned with the following problem:

$$\min f(x) - \mu \log(s) + \frac{1}{2} d_x^T D_x d_x + \frac{1}{2} d_s^T D_s d_s \quad (19a)$$

$$a(x) - s = r\mu \quad (19b)$$

$$s \geq 0 \quad (19c)$$

The KKT conditions for (??) are:

$$\nabla_x \mathcal{L}(x, y) = \nabla f(x) + D_x d_x - \nabla a(x)^T y = 0 \quad (20a)$$

$$\mathcal{C}_\mu(s, y) = Ys - \mu e = 0 \quad (20b)$$

$$\mathcal{P}_\mu(x, s) = a(x) - s - \mu r = 0 \quad (20c)$$

$$s, y \geq 0 \quad (20d)$$

Where the Lagrangian $\mathcal{L}(x, y) := f(x) - y^T a(x)$.

We combine the log barrier merit function and the complementary conditions as follows:

$$\phi(x, y) = \psi(x) + \zeta(x, y) \quad (21)$$

With:

$$\zeta(x, y) = \frac{\|\mathcal{C}(x, y)\|_\infty^3}{\mu^2}$$

We now introduces models to locally approximate these merit functions $\nabla_x \mathcal{L}(x, y)$, ψ , \mathcal{C} and ϕ respectively. To describe our approximations of a function f around the point (x, y) we use the function $\tilde{\Delta}_{(x, y)}^f(u, v)$ to denote the predicted increase in the function f at the new point $(x + u, y + v)$. Observe that we use different approximations depending on the choice of function f .

We use a typical linear approximate of $\nabla_x \mathcal{L}(x, y)$ as follows:

$$\tilde{\Delta}_{(x, y)}^{\nabla_x \mathcal{L}}(d_x, d_y) = \nabla_{x, x} \mathcal{L}(x, y) d_x + \nabla a(x) d_y \quad (22)$$

The following function $\tilde{\Delta}_{(x, y)}^\psi(u)$ is an approximation of the function $\psi(x)$ at the point (x, y) and predicts how much the function ψ changes as we change the current from x to $x + u$.

$$\tilde{\Delta}_{(x, y)}^\psi(u) = \frac{1}{2} u^T M(x, y) u + \nabla \psi(x)^T u \quad (23)$$

With:

$$M(x, y) = \nabla^2 \mathcal{L}(x, y) + \sum_i \frac{y_i}{a(x)} \nabla a(x)^T \nabla a(x) \quad (24)$$

Note that if we set $y_i = \frac{\mu}{s_i}$ then $M(x, y) = \nabla^2 \psi(x)$ and $\tilde{\Delta}_{(x, y)}^\psi$ becomes the second order taylor approximation of ψ at the point x . Thus we can think of $\tilde{\Delta}_{(x, y)}^\psi(u)$ as a primal-dual approximation of the function ψ .

We can also build a model of the $\zeta(x, y)$ as follows:

$$\tilde{\Delta}_{(x, y)}^\zeta(d_x, d_y) = \frac{\|Sy + Yd_s + Sd_y - \mu e\|_\infty^3 - \|\mathcal{C}(x, y)\|_\infty^3}{\mu^2} \quad (25)$$

With S a diagonal matrix containing entries of $a(x)$ and $d_s = \nabla a(x) d_x$. This model $\tilde{\Delta}_{(x, y)}^\zeta$ corresponds to the typical primal-dual linear model of \mathcal{C} i.e. $\mathcal{C}(x + d_x, y + d_y) \approx Sy + Yd_s + Sd_y - \mu e$.

With S and Y contain the diagonal elements of $a(x)$ and y respectively.

This allows us to approximate the change in the function ϕ at the point (x, y) as follows:

$$\tilde{\Delta}_{(x, y)}^\phi(d_x, d_y) = \tilde{\Delta}_{(x, y)}^\psi(d_x) + \tilde{\Delta}_{(x, y)}^\zeta(d_x, d_y) \quad (26)$$

We say an iterate (x, y) satisfies approximate complementary if $(x, y) \in \mathcal{Q}_\mu$ where \mathcal{Q}_μ is defined as follows:

$$\mathcal{Q}_\mu = \left\{ (x, y) \in R^n \times R^m : a(x) > 0, y > 0, \|\mathcal{C}(x, y)\|_\infty \leq \frac{\mu}{2} \right\} \quad (27)$$

We say the point (x, y) is a μ -scaled KKT point if $(x, y) \in \mathcal{T}_\mu$ where:

$$\mathcal{T}_\mu = \{(x, y) \in \mathcal{Q}_\mu : \|\nabla \mathcal{L}(x, y)\| \leq \mu(\|y\|_1 + 1)\} \quad (28)$$

In which case the algorithm terminates.

7 Algorithm

Let S, Y denote the diagonal matrices with entries of s and y respectively. We can linearize (??) at the iterate (x, y, s) as follows:

$$\begin{bmatrix} \nabla^2 \mathcal{L}(\hat{x}, \hat{y}) + D_x & -\nabla a(\hat{x})^T & 0 \\ \nabla a(\hat{x}) & 0 & -I \\ 0 & \hat{S} & \hat{Y} + D_s \end{bmatrix} \begin{bmatrix} d_x \\ d_y \\ d_s \end{bmatrix} = - \begin{bmatrix} \nabla \mathcal{L}(x, y) \\ \mathcal{P}_\mu(x, s) \\ \mathcal{C}_\mu(s, y) \end{bmatrix} \quad (29)$$

Which is equivalent to solving:

$$\begin{bmatrix} \nabla^2 \mathcal{L}(\hat{x}, \hat{y}) + \nabla a(x)^T D_s \nabla a(x) + D_x & \nabla a(\hat{x})^T \\ \nabla a(\hat{x}) & -(\hat{Y} + D_s)^{-1} \hat{S} \end{bmatrix} \begin{bmatrix} d_x \\ -d_y \end{bmatrix} = - \begin{bmatrix} \nabla \mathcal{L}(x, y) \\ \mathcal{P}_\mu(x, s) + (\hat{Y} + D_s)^{-1} \mathcal{C}_\mu(s, y) \end{bmatrix} \quad (30)$$

One can also solve this system by solving the Schur complement:

$$(\nabla^2 \mathcal{L}(\hat{x}, \hat{y}) + \nabla a(\hat{x})^T (\hat{Y} + D_s) \hat{S}^{-1} \nabla a(\hat{x}) + D_x) d_x = -\nabla \mathcal{L}(x, \mu S^{-1} e) - \nabla a(\hat{x})^T \hat{Y} \hat{S}^{-1} \mathcal{P}_\mu(x, s)$$

Observe that (??) may be singular or correspond to a direction that makes the log barrier objective worse. To rectify this problem we compute the direction as follows:

$$d_x = \arg \min_{\|u\|_2 \leq r} \tilde{\Delta}_{(x,y)}^\psi(u) \quad (31a)$$

$$d_s = \nabla a(x) d_x \quad (31b)$$

$$d_y = -S^{-1} (Y d_s + \mathcal{C}(x, y)) \quad (31c)$$

*****CAREFUL WITH SIGNS i.e. should be $d_s = -\nabla a(x) d_x$, $d_y = -S^{-1} (Y d_s + \mathcal{C}(x, y))$ *****

It is well-known from trust region literature that there exists some $\delta \in [0, \infty)$ such that:

$$(M(x, y) + \delta I) d_x = -\nabla \psi(x) \quad (32)$$

Furthermore, by re-arranging this equation we can deduce that (d_x, d_y, d_s) satisfies a perturbed version of (??):

$$\begin{bmatrix} \nabla^2 \mathcal{L}(x, y) + \delta I & -\nabla a(x)^T & 0 \\ -\nabla a(x) & 0 & I \\ 0 & S & Y \end{bmatrix} \begin{bmatrix} d_x \\ d_y \\ d_s \end{bmatrix} = - \begin{bmatrix} \nabla \mathcal{L}(x, y) \\ 0 \\ \mathcal{C}(x, y) \end{bmatrix} \quad (33)$$

Algorithm 7 Primal-dual trust region step

function PRIMAL-DUAL-TRUST-REGION(x, y, r) *****

$$d_x \in \arg \min_{\|u\| \leq r} \tilde{\Delta}_{(x,y)}^\psi(u) \quad (34a)$$

$$d_s = \nabla a(x) d_x \quad (34b)$$

$$S = \text{Diag}(a(x)) \quad (34c)$$

$$d_y = -S^{-1} (Y d_s + \mathcal{C}(x, y)) \quad (34d)$$

$$(x^+, y^+) \leftarrow (x + d_x, y + d_y)$$

return (x^+, y^+, d_x, d_y)

end function

Our complete algorithm is summarized as follows:

Algorithm 8 Primal-dual non-convex interior point algorithm

```

function NON-CONVEX-IPM( $x^1, y^1$ )
  for  $k = 1, \dots, \infty$  do
     $r \leftarrow R(y^k)$ 
    repeat
       $(x^+, y^+, d_x, d_y) \leftarrow \text{PRIMAL-DUAL-TRUST-REGION}(x^k, y^k, r)$ 
      if  $(x^+, y^+) \in \mathcal{Q}_\mu$  then
        if  $(x^+, y^+) \in \mathcal{T}_\mu$  then
          return  $(x^+, y^+)$ 
        end if
      end if
       $r \leftarrow r/2$ 
    until  $\phi(x^+) > \phi(x^k) + \frac{1}{2} \tilde{\Delta}_{(x^k, y^k)}^\phi(d_x, d_y)$ 
     $x^k \leftarrow x^+$ 
     $y^k \leftarrow y^+$ 
  end for
end function

```

8 Delta computation

Algorithm 9 Delta

```

 $\gamma_{lb} = 0, \gamma_{ub} = \delta_{\max} = \|H\|_F^2, \delta_{k-1}$   $\triangleright$  lower and upper bounds on minimum eigenvalue
Try  $\delta = 0$ , if succeeds, trial solve with this delta. If step size is small skip to trust region step.
 $\delta = \delta_{k-1}$ 
if  $\delta = 0$  then
   $\delta = \delta_{\min}$ 
end if
for  $i = 1, \dots, \infty$  do
  Break if inertia correct and update  $\gamma_{lb}$  and  $\gamma_{ub}$ .
   $\delta = \delta 100$ 
end for
Trust region
 $R = \|d_x^{k-1}\|_2$ 
for  $i = 1, \dots, \infty$  do
  Compute trust region with  $R$ 
  If trust region is too accurate increase radius size
  If step unsuccessful decrease radius size
  Prevent oscillation
end for

```
