

# Lecture 10 Nonlinear Programming Algorithms

Can Li

ChE 597: Computational Optimization  
Purdue University

# Outline of this lecture

- Equality-constrained Newton's method (only has equality constraints  $Ax = b$ )
- Interior point method
  - Barrier Method
  - Primal-Dual Interior-Point Method

## Convex objective function with linear equality constraints

Consider a problem with equality constraints, as in

$$\min_x f(x) \text{ subject to } Ax = b$$

Unconstrained convex opt.

$$\text{let } x = Fy + x_0.$$

$$x_0 \text{ be } Ax_0 = b$$

$F$  : spans of the null space of  $A$ .

$$\min_y f(Fy + x_0)$$

Equality constrained Newton's method:

Expand  $f(x)$  at point  $x_0$

$$\begin{aligned}\hat{f}(x) \approx & f(x_0) + \nabla^T f(x_0) (x - x_0) \\ & + (x - x_0)^T \nabla^2 f(x_0) (x - x_0)\end{aligned}$$

# Convex objective function with linear equality constraints

Consider a problem with equality constraints, as in

$$\min_x f(x) \text{ subject to } Ax = b$$

Options:

- Eliminating equality constraints: write  $x = Fy + x_0$ , where  $F$  spans null space of  $A$ , and  $Ax_0 = b$ . Solve in terms of  $y$  as an unconstrained optimization problem.

$$\min_x f(x) \text{ s.t. } Ax = b \quad \Leftrightarrow \quad \min_y f(Fy + x_0)$$

- Equality-constrained Newton: in many cases, this is the most straightforward option

## Equality-constrained Newton's method

In equality-constrained Newton's method, we start with  $x^{(0)}$  such that  $Ax^{(0)} = b$ . At each iteration, we solve a quadratic approximation of the problem

~~$x$  is  $x^{(0)}$  at iteration 0.~~

$$\begin{array}{ll} \min_v & \hat{f}(x+v) = f(x) + \nabla f(x)^T v + (1/2)v^T \nabla^2 f(x)v \\ \text{s.t.} & A(x+v) = b \end{array}$$

Since we kept  $Ax = b$ , the constraint is equivalent to  $Av = 0$ . We know from KKT conditions that  $v$  satisfies

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\text{nt}} \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}$$

for some dual variables  $w$ . Hence Newton direction  $v$  is again given by solving a linear system in the Hessian (albeit a bigger one). We can perform backtracking line search on the direction  $\Delta x_{\text{nt}}$

$$x^+ = x + t\Delta x_{\text{nt}}$$

where  $t$  is the appropriate step size after the line search.

# Newton's method for equality constrained minimization

**given** starting point  $x \in \text{dom } f$  with  $Ax = b$ , tolerance  $\epsilon > 0$ .

**repeat**

- Compute the Newton step and decrement  $\Delta x_{\text{nt}}, \lambda(x)$ .
- *Stopping criterion.* quit if  $\frac{\lambda^2}{2} \leq \epsilon$ .
- *Line search.* Choose step size  $t$  by backtracking line search.
- *Update.*  $x^+ := x + t \Delta x_{\text{nt}}$ .

**until** stopping criterion is met  
where the decrement is defined as

$$\lambda(x) = (\Delta x_{\text{nt}}^T \nabla^2 f(x) \Delta x_{\text{nt}})^{1/2}$$

which is the same as the standard Newton's method.

# General convex NLP

Consider the convex optimization problem

$$\begin{array}{ll}\min_x & f(x) \\ \text{subject to} & h_i(x) \leq 0, i = 1, \dots, m \\ & Ax = b\end{array}$$

We will assume that  $f, h_1, \dots, h_m$  are convex, twice differentiable, each with domain  $\mathbb{R}^n$ .



## Log barrier function

Ignoring equality constraints for now, our problem can be written as

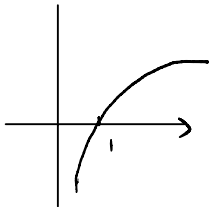
$$\begin{aligned} \min_x f(x) \\ \text{s.t. } h_i(x) \leq 0 \quad \forall i \end{aligned} \Leftrightarrow \min_x f(x) + \sum_{i=1}^m I_-(h_i(x)) \quad I_- = \begin{cases} 0 & h_i(x) \leq 0 \\ +\infty & h_i(x) > 0 \end{cases}$$

Where  $I_-$  represents the indicator function (discontinuous)

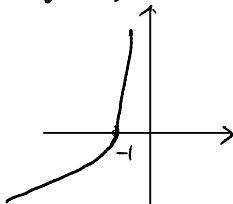
$$I_-(u) = \begin{cases} 0 & u \leq 0 \\ +\infty & u > 0 \end{cases}$$



$\log(x)$



$-\log(-x)$



## Log barrier function

Ignoring equality constraints for now, our problem can be written as

$$\min_x f(x) + \sum_{i=1}^m l_-(h_i(x))$$

Where  $l_-$  represents the indicator function (discontinuous)

$$l_-(u) = \begin{cases} 0 & u \leq 0 \\ +\infty & u > 0 \end{cases}$$

We can approximate the sum of indicators by the log barrier:

$$\min_x f(x) - \frac{1}{t} \sum_{i=1}^m \log(-h_i(x))$$

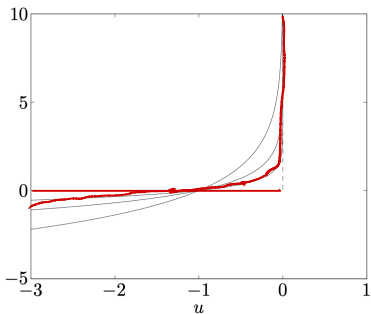
where  $t > 0$  is a large number

This approximation is more accurate for larger  $t$ . But for any value of  $t$ , the log barrier approaches  $\infty$  if any  $h_i(x) \rightarrow 0$

## Interpretation of log barrier function

$$\min_x f(x) - \frac{1}{t} \sum_{i=1}^m \log(-h_i(x))$$

where  $t > 0$  is a large number



The dashed lines show the function  $I_-(u)$ , and the solid curves show  $\hat{I}_-(u) = -(1/t) \log(-u)$ , for  $t = 0.5, 1, 2$ . The curve for  $t = 2$  gives the best approximation.

## Log barrier calculus

For the log barrier function

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

$$V V^T$$

$$V \times 6 \mathbb{R}^n$$

$$x^T V V^T x$$

$$= \|V^T x\|_2^2$$

we have for its gradient:

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

and for its Hessian:

$$\nabla^2 \phi(x) = \sum_{i=1}^m \frac{1}{h_i(x)^2} \nabla h_i(x) \nabla h_i(x)^T - \sum_{i=1}^m \frac{1}{h_i(x)} \nabla^2 h_i(x)$$

*Handwritten notes: The first term is underlined with a red squiggle. The second term is circled in red. Above the circle is a red "> 0" and below it is a red "< 0".*

The Hessian is always PSD for  $h_i(x) < 0$ . Thus, the barrier problem is convex.

## Central path

$$\min f(x) + \frac{1}{t} \phi(x)$$

Consider barrier problem:

$$\begin{array}{ll} \min_x & tf(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array}$$

The central path is defined by solution  $x^*(t)$  with respect to  $t > 0$

- This is a convex optimization problem with one equality constraint. It can be solved by the equality-constrained Newton's method.
- Hope is that, as  $t \rightarrow \infty$ , we will have  $x^*(t) \rightarrow x^*$ , solution to our original problem
- Why don't we just set  $t$  to be some huge value, and solve the above problem? Directly seek solution at end of central path?
- Problem is that this is seriously inefficient in practice
- Much more efficient to traverse the central path, as we will see

## How to estimate the duality gap of the original problem?

- The barrier problem give us a strictly feasible primal solution  $x^*(t)$  to the original problem.
- If we would like to estimate how far the  $f(x^*(t))$  is from the optimum, we will need the dual variables of the original problem because we need that to estimate duality gap.
- Idea: use the KKT conditions of the barrier problem to construct a dual feasible solution to the original problem.

# KKT conditions of barrier and the original problem

KKT conditions of the barrier problem.

$$\begin{array}{ll} \min & t f(x) + \phi(x) \\ \text{s.t.} & Ax = b \end{array}$$

(stationarity)

$$\nabla f(x^*(t)) - \sum_{i=1}^m \frac{1}{h_i(x^*(t))} \nabla h_i(x^*(t)) + A^T w = 0,$$

(primal feasibility)

$$Ax^*(t) = b, \quad h_i(x^*(t)) < 0, \quad i = 1, \dots, m$$

$\min f(x)$   
 $\text{s.t. } h_i(x) \leq 0$   
 $Ax = b$

for some  $w \in \mathbb{R}^m$ .

Since we know  $x^*(t)$  is optimal, it satisfies the KKT conditions of the barrier problem.

**Claim** Given  $x^*(t)$  and corresponding  $w$ , we define

$$u_i^*(t) = -\frac{1}{th_i(x^*(t))}, \quad i = 1, \dots, m, \quad v^*(t) = w/t$$


$u^*(t)$  and  $v^*(t)$  are dual feasible for the original problem, i.e.,  $g(u^*(t), v^*(t)) \neq -\infty$ , where  $u_i^*(t)$  is the dual variable for  $h_i(x) \leq 0$ ,  $v^*(t)$  are the dual variable for  $Ax = b$

**Proof** The Lagrangian of the original problem is

$$L(x, u, v) = f(x) + \sum_{i=1}^m u_i h_i(x) + v^T (Ax - b)$$

- Note that  $u_i^*(t) > 0$  since  $h_i(x^*(t)) < 0$  for all  $i = 1, \dots, m$  (the signs are correct)
- The next step is to check whether  $g(u, v)$  is finite at  $(u^*(t), v^*(t))$ .

$$g((u^*(t), v^*(t))) = \min_x f(x) + \sum_{i=1}^m u_i^*(t) h_i(x) + v^*(t)^T (Ax - b)$$

$x^*(t)$  minimize 

- By definition (from the KKT condition of the barrier problem)

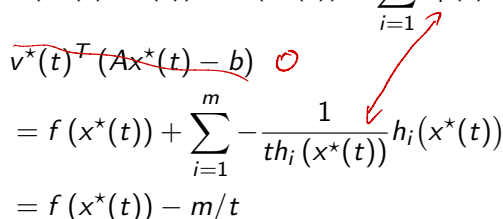
$$\nabla f(x^*(t)) + \sum_{i=1}^m u_i(x^*(t)) \nabla h_i(x^*(t)) + A^T v^*(t) = 0$$

That is,  $x^*(t)$  minimizes Lagrangian  $L(x, u^*(t), v^*(t))$  over  $x$ ,  
so  $g(u^*(t), v^*(t)) > -\infty$



## Duality gap

This allows us to bound suboptimality of  $f(x^*(t))$ , with respect to original problem, via the duality gap. We compute

$$\begin{aligned} g(u^*(t), v^*(t)) &= f(x^*(t)) + \sum_{i=1}^m u_i^*(t) h_i(x^*(t)) + \\ &\quad \cancel{v^*(t)^T (Ax^*(t) - b)} \quad \circ \\ &= f(x^*(t)) + \sum_{i=1}^m -\frac{1}{th_i(x^*(t))} h_i(x^*(t)) \\ &= f(x^*(t)) - m/t \end{aligned}$$


That is, we know that  $f(x^*(t)) - f^* \leq m/t$

This will be very useful as a stopping criterion; it also confirms the hope that  $x^*(t) \rightarrow x^*$  as  $t \rightarrow \infty$

## Interpretation: Perturbed KKT conditions

We can now reinterpret central path  $(x^*(t), u^*(t), v^*(t))$  as solving the perturbed KKT conditions:

$$\begin{aligned}\nabla f(x) + \sum_{i=1}^m u_i \nabla h_i(x) + A^T v &= 0 \\ u_i \cdot h_i(x) &= -1/t, \quad i = 1, \dots, m \\ h_i(x) &\leq 0, \quad i = 1, \dots, m, \quad Ax = b \\ u_i &\geq 0, \quad i = 1, \dots, m\end{aligned}$$

Only difference between these and actual KKT conditions for our original problem is second line: these are replaced by

$$u_i \cdot h_i(x) = 0, \quad i = 1, \dots, m$$

i.e., complementary slackness, in actual KKT conditions

# Barrier Method

The barrier method solves a sequence of problems

$$\begin{array}{ll} \min_x & tf(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array}$$

for increasing values of  $t > 0$ , until duality gap satisfies  $m/t \leq \epsilon$ .

1. We fix  $t^{(0)} > 0, \mu > 1$ . We use Newton to compute  $x^{(0)} = x^*(t)$ , solution to barrier problem at  $t = t^{(0)}$ .
2. For  $k = 1, 2, 3, \dots$ , solve the barrier problem at  $t = t^{(k)}$ , using Newton initialized at  $x^{(k-1)}$ , to yield  $x^{(k)} = x^*(t)$
3. Stop if  $m/t \leq \epsilon$ , else update  $t^{(k+1)} = \mu t$

Step 2 is called a centering step (since it brings  $x^{(k)}$  onto the central path)

## Choice of parameters

- Choice of  $\mu$  : if  $\mu$  is too small, then many outer iterations might be needed; if  $\mu$  is too big, then Newton's method (each centering step) might take many iterations
- Choice of  $t^{(0)}$  : if  $t^{(0)}$  is too small, then many outer iterations might be needed; if  $t^{(0)}$  is too big, then the first Newton solve (first centering step) might require many iterations

Fortunately, the performance of the barrier method is often quite robust to the choice of  $\mu$  and  $t^{(0)}$  in practice (However, note that the appropriate range for these parameters is scale dependent)

# Barrier versus primal-dual interior point method

## Overview:

- Both can be motivated in terms of perturbed KKT conditions
- Primal-dual interior-point methods take one Newton step, and move on (no separate inner and outer loops)
- Primal-dual interior-point iterates are not necessarily feasible
- Primal-dual interior-point methods are often more efficient, as they can exhibit better than linear convergence
- Primal-dual interior-point methods are less intuitive

## Interpretation: Perturbed KKT conditions

$$F(x) = 0$$

We can now reinterpret central path  $(x^*(t), u^*(t), v^*(t))$  as solving the perturbed KKT conditions:

$$\nabla f(x) + \sum_{i=1}^m u_i \nabla h_i(x) + A^T v = 0$$

$$u_i \cdot h_i(x) = -1/t, \quad i = 1, \dots, m$$

$$h_i(x) \leq 0, \quad i = 1, \dots, m, \quad Ax = b$$

$$u_i \geq 0, \quad i = 1, \dots, m$$

Only difference between these and actual KKT conditions for our original problem is second line: these are replaced by

$$u_i \cdot h_i(x) = 0, \quad i = 1, \dots, m$$

i.e., complementary slackness, in actual KKT conditions

## Perturbed KKT as nonlinear system

Can view this as a nonlinear system of equations, written as

$$r(x, u, v) = \begin{pmatrix} \nabla f(x) + Dh(x)^T u + A^T v \\ -\text{diag}(u)h(x) - (1/t)1 \\ Ax - b \end{pmatrix} = 0$$

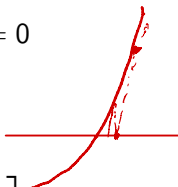
where

$$h(x) = \begin{pmatrix} h_1(x) \\ \vdots \\ h_m(x) \end{pmatrix}, \quad Dh(x) = \begin{bmatrix} \nabla h_1(x)^T \\ \vdots \\ \nabla h_m(x)^T \end{bmatrix}$$

Newton's method, recall, is generally a root-finder for a nonlinear system  $F(y) = 0$ . Approximating  $F(y + \Delta y) \approx F(y) + DF(y)\Delta y = 0$  leads to

$$\Delta y = -(DF(y))^{-1}F(y)$$

What happens if we apply this to  $r(x, u, v) = 0$  above?



## Newton on perturbed KKT, v1

Approach 1: from middle equation (relaxed comp slackness), note that  $u_i = -1/(th_i(x))$ ,  $i = 1, \dots, m$ . So after eliminating  $u$ , we get

$$\mu_i h_i(x) = -\frac{1}{t}$$

$$r(x, v) = \begin{pmatrix} \nabla f(x) + \sum_{i=1}^m \left( -\frac{1}{th_i(x)} \right) \nabla h_i(x) + A^T v \\ Ax - b \end{pmatrix} = 0$$

Thus the Newton root-finding update  $(\Delta x, \Delta v)$  is determined by

$$\begin{bmatrix} H_{\text{bar}}(x) & A^T \\ A & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta v \end{pmatrix} = -r(x, v)$$

where  $H_{\text{bar}}(x) =$

$$\nabla^2 f(x) + \sum_{i=1}^m \frac{1}{th_i(x)^2} \nabla h_i(x) \nabla h_i(x)^T + \sum_{i=1}^m \left( -\frac{1}{th_i(x)} \right) \nabla^2 h_i(x)$$

This is just the KKT system solved by one iteration of Newton's method for minimizing the barrier problem



## Newton on perturbed KKT, v2

Approach 2: directly apply Newton root-finding update, without eliminating  $u$ . Introduce notation

$$r_{\text{dual}} = \nabla f(x) + Dh(x)^T u + A^T v$$

$$r_{\text{cent}} = -\text{diag}(u)h(x) - (1/t)t$$

$$r_{\text{prim}} = Ax - b$$

called the dual, central, and primal residuals at  $y = (x, u, v)$ . Now root-finding update  $\Delta y = (\Delta x, \Delta u, \Delta v)$  is given by

$$\begin{bmatrix} H_{\text{pd}}(x) & Dh(x)^T & A^T \\ -\text{diag}(u)Dh(x) & -\text{diag}(h(x)) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta u \\ \Delta v \end{pmatrix} = - \begin{pmatrix} r_{\text{dual}} \\ r_{\text{cent}} \\ r_{\text{prim}} \end{pmatrix}$$

where  $H_{\text{pd}}(x) = \nabla^2 f(x) + \sum_{i=1}^m u_i \nabla^2 h_i(x)$

## Surrogate duality gap

For barrier method, we have simple duality gap:  $m/t$ , since we set  $u_i = -1/(th_i(x))$ ,  $i = 1, \dots, m$  and saw this was dual feasible  
For primal-dual interior-point method, we can construct surrogate duality gap:

$$\eta = -h(x)^T u = -\sum_{i=1}^m u_i h_i(x)$$

This would be a real duality gap if we had feasible points, i.e.,  $r_{\text{prim}} = 0$  and  $r_{\text{dual}} = 0$ , but we don't, so it's not  
What value of parameter  $t$  does this correspond to in perturbed KKT conditions? This is  $t = m/\eta$

# Primal-dual Interior-point Method

Putting it all together, we now have our primal-dual interior-point method.

- Start with  $x^{(0)}$  such that  $h_i(x^{(0)}) < 0$ ,  $i = 1, \dots, m$ , and  $u^{(0)} > 0$ ,  $v^{(0)}$ . Define  $\eta^{(0)} = -h(x^{(0)})^T u^{(0)}$ . We fix  $\mu > 1$ , repeat for  $k = 1, 2, 3, \dots$ 
  - Define  $t = \mu m / \eta^{(k-1)}$
  - Compute primal-dual update direction  $\Delta y$
  - Use backtracking to determine step size  $s$
  - Update  $y^{(k)} = y^{(k-1)} + s \cdot \Delta y$
  - Compute  $\eta^{(k)} = -h(x^{(k)})^T u^{(k)}$
  - Stop if  $\eta^{(k)} \leq \varepsilon$  and  $(\|r_{\text{prim}}\|^2 + \|r_{\text{dual}}\|^2)^{1/2} \leq \varepsilon$

Note that we stop based on surrogate duality gap, and approximate feasibility. (Line search maintains  $h_i(x) < 0$ ,  $u_i > 0$ ,  $i = 1, \dots, m$ )

## Backtracking line search

At each step, must ensure we arrive at  $y^+ = y + s\Delta y$ , i.e.,

$$x^+ = x + s\Delta x, \quad u^+ = u + s\Delta u, \quad v^+ = v + s\Delta v$$

that maintains both  $h_i(x) < 0$ , and  $u_i > 0, i = 1, \dots, m$

A multi-stage backtracking line search for this purpose: start with largest step size  $s_{\max} \leq 1$  that makes  $u + s\Delta u \geq 0$  :

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

Then, with parameters  $\alpha, \beta \in (0, 1)$ , we set  $s = 0.999s_{\max}$ , and

- Let  $s = \beta s$ , until  $h_i(x^+) < 0, i = 1, \dots, m$
- Let  $s = \beta s$ , until  $\|r(x^+, u^+, v^+)\|_2 \leq (1 - \alpha s)\|r(x, u, v)\|_2$

## Software

Gurobi  
LP

- ① primal simplex
- ② dual simplex
- ③ barrier algorithm, with crossover
- ④ push to an extreme

- Gurobi/Cplex has barrier algorithm for linear programs. point.  
Barrier algorithm has no guarantee to terminate at an extreme point.
- Interior point solvers: IPOPT (open-source), Mosek, Knitro

## References

- Convex optimization notes by Ryan Tibshirani  
<https://www.stat.cmu.edu/~ryantibs/convexopt/>
- Boyd, S. P., & Vandenberghe, L. (2004). Convex optimization. Cambridge university press.
- Lemaréchal, C., & Oustry, F. (1999). Semidefinite relaxations and Lagrangian duality with application to combinatorial optimization (Doctoral dissertation, INRIA).