Numerical Optimisation
Constraint optimisation:
Interior point methods

f.rullan@cs.uc

Department of Computer Sc
Centre for Medical Image Com
Centre for Inverse Problems
University College London

Lecture 15

Convex constraint optimization problem

$$\min_{x \in \mathcal{D} \subset \mathbb{R}^n} f(x) \qquad \text{(COP)}$$
$$\text{subject to} \quad f_i(x) \quad 0, \quad i = 1, \dots, m,$$

wher

- $f : \mathcal{D} \to \mathbb{R}$ is convex, twice continuou function, $\mathcal{D} \subset \mathbb{R}^n$ is convex

- $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots m$ are conv differentiable functions
- $A \in \mathbb{R}^{p \times n}$ with rank $A = p < n$.

We assume that

- (COP) is solvable i.e. an optimal $x^\star$ exists, and we denote the optimal value as $p^\star = f(x^\star)$.

- (COP) is strictly feasible i.e. there exists $x \in \mathcal{D}$ that satisfies $Ax = b$ and $f_i(x) < 0$ for $i = 1, \dots, m$. This means that Slater's constraint qualification holds, thus there exists dual

tisfy the

$$\nabla f(x^\star) + \sum_{i=1} \lambda_i^\star \nabla f_i(x^\star) + A^{\mathrm{T}} \quad {}^\star \qquad \text{KT)}$$

$$f_i(x \ ) \leq 0, \quad i = 1, \dots, m,$$

$$\lambda^\star \geq 0,$$

$$\lambda_i^\star f_i(x^\star) = 0, \quad i = 1, \dots m.$$

Interior point methods solve either

- the problem (COP) by applying Newton's method to a
-

We will consider the *barrier method* a
interior-point method.

Rewrite (COP) making the inequality constraints implicit

$$\min_{x \in \mathcal{D} \subset \mathbb{R}^n} f(x) + \sum_{i=1}^{m} I_-(f_i(x))$$

wher
reals

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

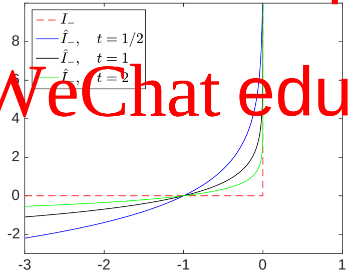
I_- is non-differentiable thus we need a smooth approximation before Newton method can be applied.

Approximate I_- with a smooth *logarithmic barrier*

$$\hat{I}_-(u) = -1/t \log(-u), \quad \operatorname{dom} \hat{I}_- = [-\infty, 0),$$

where $t > 0$ is a parameter that sets the accuracy of the approximation.

- Like I_-, $\hat{I}_-$ is convex nondecreasing and by convention ∞ for $u > 0$.
- ∞



Substituting $\hat{I}_-$ for I_- yields an approximation

$$\min_{x \in \mathcal{D} \subset \mathbb{R}^n} \quad f(x) + \sum_{i=1}^{m} -1/t \log(-f_i(x))$$
$$\text{subject to} \quad Ax = b.$$

The objective function is convex since $-1/t \log(-u)$ is increasing in u, and differentiable, thus Newton's method can be applied.

Log

$$\phi(x) = -\sum_{i=1}^{m} \log(-f_i(x)), \quad \operatorname{dom}\phi = \{x \in \mathbb{R}^n : f_i(x) < 0, \ i = 1, \dots, m\}$$

Gradient and Hessian of ϕ

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

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

Consider the equivalent problem

$$\min_{x \in \mathcal{D} \subset \mathbb{R}^n} \quad tf(x) + \phi(x) \qquad \text{(CENT)}$$

subject to $Ax = b$.

We assume that (CENT) has a unique solution for each $t > 0$, and deno[illegible]

The s[illegible] e

point[illegible]

and sufficient *centrality conditions*:

$x^\star(t)$ is strictly feasible i.e. satisfies

$$Ax^\star(t) = b, \quad f_i(x^\star(t)) < [illegible]$$

and there exists a $\hat{\nu} \in \mathbb{R}^p$ such that

$$0 = t\nabla f(x^\star(t)) + \nabla\phi(x^\star(t)) + A^{\mathrm{T}}\hat{\nu} \qquad \text{(CENT:COND)}$$

$$= t\nabla f(x^\star(t)) + \sum_{i=1}^{m} \frac{1}{-f_i(x^\star(t))} \nabla f(x^\star(t)) + A^{\mathrm{T}}\hat{\nu}$$

$$\min_{x \in \mathbb{R}^n} \quad c^{\mathrm{T}} x$$
$$\text{subject to} \quad Ax \leq b.$$

The logarithmic barrier:

$$\phi(x) = -\sum_{i=1}^{m} \log(b_i - a_i^{\mathrm{T}} x), \quad \operatorname{dom}\phi = \{x : Ax < b\}.$$

The g

$$\nabla\phi(x) = \sum_{i=1}^{m} \frac{1}{b_i - a_i^{\mathrm{T}} x} a_i, \quad \nabla^2\phi(x \ldots$$

Since x is strictly feasible, we have b_i ...
nonsingular iff A has rank n.
The centrality condition (CENT:COND): $(\nabla\phi(x^\star(t)) \parallel -c)$

$$tc + \sum_{i=1}^{m} \frac{1}{b_i - a_i^{\mathrm{T}} x} a_i = 0.$$

Figure: Boyd Vandenberghe Fig. 11.2

Dual points from central path

Claim: Every point on central path yields a dual feasible point and hence a lower bound on $p^\star$. More precisely, the pair

$$\lambda_i^\star(t) = \frac{1}{t f_i(x^\star(t))}, \quad i = 1, \dots, m, \quad \nu^\star(t) = \nu/t$$

is dual f

Pro $) < 0$

and from optimality conditions (CENT:CON

$x^\star(t)$ minimises the Lagrangian

$$\mathcal{L}(x, \lambda, \nu) = f(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \nu^{\mathrm{T}}(Ax - b)$$

for $\lambda = \lambda^\star(t), \nu = \nu^\star(t)$.

This means that $\lambda^\star(t), \nu^\star(t)$ are dual feasible, the dual function is finite and

$$g(\lambda^\star(t), \nu^\star(t)) = f(x^\star(t)) + \sum_{i=1}^{m} \underbrace{\lambda_i^\star(t)}_{=-\frac{1}{tf_i(x^\star(t))}} f_i(x^\star(t))$$

thus $x^\star(t)$ is not more than m/t subo

$$f(x^\star(t)) - p^\star$$

and $x^\star(t)$ converges to an optimal point as $t \to \infty$.

We can interpret the central path conditions as a continuous deformation of (KKT). A point x is equal to $x^\star(t)$ iff there exists λ, ν such that

m

:CENT)

$$f_i(x) \quad 0, \quad i = 1, \dots, m,$$

$$\lambda$$

$$-\lambda_i f_i(x) = 1$$

The only difference to (KKT) is the complementarity condition $-\lambda_i f_i(x) = 1/t$. For large t, $x^\star(t), \lambda^\star(t), \nu^\star(t)$ almost satisfy the KKT conditions.

The Newton step for the centering problem (CENT) (linearly equality constraint problem) reads

$$\left[\quad {}^{2} \quad {}^{2} \quad {}^{T} \quad \ldots \quad x) \right].$$

Her

We can interpret these Newton step for (CENT) a directly solving the modified (KKT:CENT) in a p

Newton for modified KKT (KKT:CENT)

First, eliminate λ using $\lambda_i = -1/(tf_i(x))$ from the (KKT:CENT) system

$$\nabla f(x) + \sum_{i=1}^{m} \frac{1}{-tf_i(x)} \nabla f_i(x) + A^{\mathrm{T}}\nu = 0, \quad Ax = b.$$

To fin

above, we form the Taylor expansion for the nonli

$$\begin{aligned}&\nabla f(x+v) + \frac{1}{t}\nabla\phi(x+v)\\ &\approx \underbrace{\nabla f(x) + \frac{1}{t}\nabla\phi(x)}_{=:g} + \underbrace{\left(\nabla^2 f(x) + \frac{1}{t}\nabla^2\phi(x)\right) v}_{=:Hv}.\end{aligned}$$

Replace the nonlinear term with this linear approximation

$$Hv + A^{\mathrm{T}}\nu = -g, \quad Av = 0$$

and observe that the Newton step $\Delta x_n, \nu_n$ for (CENT) satisfies

Com

$$v = \Delta x_n, \quad \nu = (1$$

This shows that the Newton step for the centring pr can be interpreted (after scaling of the dual variable) as the Newton step for solving the modified (KKT:CENT) system.

Require: Strictly feasible $x^s := x^{(0)}$, $t := t^{(0)} > 0$, $\mu > 1$
Require: Tolerance $\epsilon > 0$

loop

  if $m/t < \epsilon$ **then**
    break {stopping criterium ϵ-s
  end if
  Increase $t = \mu t$
end loop

- **Centering step**: can be solved by any methods for linearly constraint minimisation, in particular Newton method. Exact solve is not necessary since the central path has no significance beyond that it leads to the solution of the original

  difference in cost between the exact and goo
  marginal, few Newton steps.

- **Choice of μ:** trade of between the num
  (Newton) iterations, but it is not very critical. Values around 10-20 seem to work well.

- **Choice of $t^{(0)}$:** Trade of between the number of inner iterations in the first step and number of outer iterations.
  - Choose so that $m/t^{(0)} \approx f(x^{(0)}) - p^\star$. For instance if a dual feasible point λ, ν is known with the duality gap $\eta = f(x^{(0)}) - g(\lambda, \nu)$, then we can take $t^{(0)} = m/\eta$ (the first centering step will compute a pair with the same duality gap as the initial primal and dual feasible points).
  - Choose $t^{(0)}$ as a minimiser of

problem for $t^{(0)}, \nu$).

- **Infeasible Newton method:** for c
  $x^{(0)} \in \mathcal{D}$, $f_i(x^{(0)}) < 0, i = 1, \ldots,$
  $Ax^{(0)} = b$. Assuming the centering problem is strictly feasible, a full Newton step is taken at some point during the first centering step and thereafter the iterates are primal feasible and the algorithm coincides with the standard barrier method.

The barrier method requires a strictly feasible point $x^{(0)}$. When such a point is not known, the barrier method is preceded by a preliminary stage called *phase I* to compute a strictly feasible point (or to fi

Cons

$$f_i(x) \leq 0, \; i = 1, \dots, m, \quad Ax = b \qquad \text{(FEAS)}$$

Assume we have a point $x^{(0)} \in \prod_{i=1}^m \operatorname{dom} f_i$ and $Ax^{(0)} = b$ i.e. the inequalities are possibly not satisfied at $x^{(0)}$.

Goal: find a strictly feasible solution of equalities and inequalities:

$$\min \quad s \qquad \text{(PH1:MAX)}$$

s: bou

goal is to drive this maximum below 0.

The problem (PH1:MAX) is always strictly feas
initialise with $x = x^{(0)}$ and for s with an
$\max_{i=1,\dots,m} f_i(x^{(0)})$ and apply the barrier method.

Let $p_I^\star$ denote the optimal value for (PH1:MAX).

- $p_I^\star < 0$: (FEAS) has a strictly feasible solution.
  If (x, s) is feasible for (PH1:MAX) with $s < 0$, then x satisfies
  $f_i(x) < 0$.

  We do not need to solve (PH1:MAX) with high accuracy, we

-
  can terminate when a dual feasible point is found with
  positive objective, which proves that

- If $p_I^\star = 0$ and the minimum is attained at n

  the set of inequalities is feasible, but not strict
  $p^\star = 0$ and the minimum is not attained, the inequalities are
  infeasible.

$$
\begin{aligned}
\min \quad & \mathbf{1}^{\mathrm{T}} s && \text{(PH1:SUM)}\\
\text{subject to} \quad & f_i(x) \le s, \quad i = 1, \ldots, m &&\\
& Ax = b &&\\
& s \ge 0 &&
\end{aligned}
$$

- hus
- The optimal value is 0 and achieved iff the origi equalities and inequalities is feasible.
- When the system of equalities and inequali often the solution violates only a small numb i.e. we identified a large feasible subset. This is more informative than finding that m inequalities together are mutually infeasible.

Termination near phase II central path

Assume $x^{(0)} \in \mathcal{D} \cap \prod_{i=1}^{m} \operatorname{dom} f_i$ with $Ax^{(0)} = b$.

Modified phase I optimisation problem

$$Ax = b$$

with $M > \max\{f(x^{(0)}), p^\star\}$.

Central path for this modified problem intersects the central path for the original problem (COP).

We expresse (COP) in equivalent form

$$
\begin{aligned}
\min \quad & f(x) \\
\text{subject to} \quad & f_i(x) \leq s, \quad i = 1, \ldots, m \\
& Ax = b, \quad s = 0.
\end{aligned}
$$

Star

solve

$$
\begin{aligned}
\min \quad & tf(x) - \sum_{i=1} \log(s \quad f_i(x)) \\
\text{subject to} \quad & Ax = b \quad s
\end{aligned}
$$

which can be initialised with any $x \in \mathcal{D}$

i i

infeasibility

Provided the problem is strictly feasible, the infeasible start Newton will eventually take an undamped step an thereafter we will have $s = 0$ i.e. x strictly feasible.

The same trick can be applied if a point in $\mathcal{D} \cap \prod_{i=1}^{m} \operatorname{dom} f_i$ (domain of the function and inequality constraints) is not known.

Apply infeasible Newton to

subject to $\quad Ax = b, \; s = 0, \; z_0$

with initialisation $z_i : x + z_i \in \operatorname{dom} f_i$.

Disadvantage: no good stopping criterion for infeasible problems; the residual simply fails to converge to 0.

- Typically the cost of solving a set of convex inequalities and linear equalities using the barrier method is modest, and approximately constant, as long as the problem is not very close to the boundary between feasibility and infeasibility. When the problem is very close to the boundary, the number of Newton steps required to find a strictly feasible point or

cost becomes infinite.

- Typically the infeasible start Newton met provided the inequalities are feasible, and n boundary between feasible and infeasible. But when the feasible set is just barely nonempty, a phase I method is far better. Another advantage of the phase I method is that it gracefully handles the infeasible case; the infeasible start Newton method, in contrast, simply fails to converge.

Primal-dual interior point method is similar to barrier method with key differences:

- There is only one loop or iteration, i.e., there is no distinction between inner and outer iterations as in the barrier method.

- are obtained from Newton's method, applied to modified KKT equations (i.e., the optimality condition
barrier centering problem). The primal-
are similar to, but not quite the same as, the sea
that arise in the barrier method.

- In a primal-dual interior-point method, the primal and dual iterates are not necessarily feasible.

As in barrier method we start from (KKT:CENT) which we rewrite in the form

$$0 = r_t(x,\lambda,\nu) = \begin{bmatrix} \nabla f(x) + J(x)^{\mathrm{T}}\lambda + A^{\mathrm{T}}\nu \\ -\operatorname{diag}(\lambda)F(x) - (1/t)\mathbf{1} \\ Ax - b \end{bmatrix} =: \begin{bmatrix} r_{\mathrm{dual}} \\ r_{\mathrm{cent}} \\ r_{\mathrm{pri}} \end{bmatrix},$$

with

$$F(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix}, \quad J(x) = Df(x) = \begin{bmatrix} \nabla f_1(x)^{\mathrm{T}} \\ \vdots \\ \nabla f_m(x)^{\mathrm{T}} \end{bmatrix}.$$

If x, λ, ν satisfy $r_t(x,\lambda,\nu) = 0$ (and $f_i(x) < 0$), then $x = x^\star(t), \lambda = \lambda^\star(t), \nu = \nu^\star(t)$. In particular, x is primal feasible, and λ, ν are dual feasible, with duality gap m/t.

Newton step for solution of $r_t(x, \lambda, \nu) = 0$ at $y = (x, \lambda, \nu)$ a primal-dual strictly feasible point $F(x) < 0, \lambda > 0$.

Difference to barrier method: we do not eliminate λ before taking the Newton step

wher

Written in terms of x, λ, ν:

$$\begin{bmatrix} \nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 f_i(x) & J(x) & & \\ -\operatorname{diag}(\lambda) J(x) & -\operatorname{diag}(& & \\ A & 0 & 0 & \Delta\nu \end{bmatrix} \quad \begin{bmatrix} r_{\text{dual}} \\ r_{\text{cent}} \\ r_{\text{prim}} \end{bmatrix}.$$

(PD:N)

Eliminate $\Delta\lambda_{\mathrm{pd}}$ from (PD:N):

From the second block

$$\Delta\lambda_{\mathrm{pd}} = -\mathrm{diag}(f(x))^{-1}\mathrm{diag}(\lambda) J(x)\Delta x_{\mathrm{pd}} + \mathrm{diag}(f(x))^{-1} r_{\mathrm{cent}}$$

and su

$$\begin{bmatrix} H_{\mathrm{p}} \\ A \end{bmatrix}$$

$$= -\begin{bmatrix} \nabla f(x) + \frac{1}{t} \end{bmatrix},$$

where

$$H_{\mathrm{pd}} = \nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 f_i(x) + \sum_{i=1}^{m} \frac{\lambda_i}{-f_i(x)} \nabla f_i(x) \nabla f_i(x)^{\mathrm{T}}.$$

Compare to the Newton step in the barrier method (in the infeasible form)

$$\begin{bmatrix} H_{\mathrm{bar}} & A^{\mathrm{T}} \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{bar}} \\ \nu_{\mathrm{bar}} \end{bmatrix} = - \begin{bmatrix} t\nabla f(x) + \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x) \\ r_{\mathrm{pri}} \end{bmatrix},$$

wher

H_{ba} ... $f_i(x)^{\mathrm{T}}$.

Multiplying first block by $1/t$ and chang

$\Delta\nu_{\mathrm{bar}} = (1/t)\nu_{\mathrm{bar}} - \nu$

$$\begin{bmatrix} \frac{1}{t} H_{\mathrm{bar}} & A^{\mathrm{T}} \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{bar}} \\ \Delta\nu_{\mathrm{bar}} \end{bmatrix} = - \begin{bmatrix} \nabla f(x) + \frac{1}{t} \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x) + A^{\mathrm{T}}\nu \\ r_{\mathrm{pri}} \end{bmatrix},$$

The right hand sides are identical.

The only difference are

$$H_{\mathrm{p}} \quad 2 \quad \overset{m}{} \quad 2 \quad \overset{m}{} \quad \frac{\lambda_i}{\quad} \quad x)^{\mathrm{T}}.$$

$$\frac{1}{t} H_{\mathrm{b}} \quad \sum_{i=1} \frac{\quad}{i} \quad \sum_{i=1} \frac{\quad}{i} \quad)\nabla f_i(x)^{\mathrm{T}}.$$

When x, λ, ν satisfy $-f_i(x)\lambda_i = 1/t$, th
hence directions) coincide.

- In the primal-dual interior point methods, the iterates $x^{(k)}, \lambda^{(k)}, \nu^{(k)}$ are not necessarily feasible, except in the limit as the algorithm converges.

- Hence, cannot easily evaluate duality gap $\eta^{(k)}$ in the $k$the

- x that satisfied $F(x) < 0$ and $\lambda \geq 0$ as

$$\eta(x, \lambda) = -f$$

- The surrogate gap is the duality gap if x were primal feasible and λ, μ were dual feasible i.e. if $r_{\mathrm{prim}} = 0$, $r_{\mathrm{dual}} = 0$. Note that value of t corresponds to the surrogate duality gap $\eta \approx m/t \to t = m/\eta$.

Require: x that satisfies $F(x) < 0$, $\lambda > 0$
Require: $\mu > 1$
Req

1: **r**

2:

3: $_{\text{pd}}$

4: Line search: determine step length
$y := y + s\Delta y_{\text{pd}}$

5: **until** $\|r_{\text{prim}}\| \leq \epsilon_{\text{feas}}, \|r_{\text{dual}}\| \leq \epsilon_{\text{fea}}$

- The parameter t is set to a factor $\mu m/\eta$, which is the value of t associated with the current surrogate duality gap η. If x, λ, ν were central, with parameter t (and therefore with duality gap m/t), then we would increase t by the factor μ (as

-

- The primal-dual interior-point algorith is primal feasible and λ, ν are dual feas ϵ_{feas}) and the surrogate gap is smaller than the tolerance ϵ. Since the primal-dual interior-point method often has faster than linear convergence, it is common to choose $\epsilon_{\text{feas}}, \epsilon$ small.

- The line search in the primal-dual interior point method is a standard backtracking line search, based on the norm of the residual, and modified to ensure that $\lambda > 0$ and $F(x) < 0$.
- Start with $s_{\max} = \sup\{s \in [0,1] : \lambda + s\Delta\lambda \geq 0\}$, multiply by $\rho \in (0,1)$ until $F(x + s\Delta x_{\mathrm{pd}}) < 0$. Continue multiplying until we have

$$, \lambda, \nu)\|.$$

0.8.

- One iteration of the primal-dual interior-p
same as one step of the infeasible Newton met
solving $r_t(x, \lambda, \nu) = 0$, but modified to ens
$F(x) < 0$ (or, equivalently, with $\mathrm{dom}\, r_t$ restricted to $\lambda > 0$ and $F(x) < 0$). The same arguments used in the proof of convergence of the infeasible start Newton method show that the line search for the primal-dual method always terminates in a finite number of steps.