Numerical Optimisation
Constraint optimisation: Interior point methods

Marta M. Betcke
m.betcke@ucl.ac.uk
Kiko Rullan
f.rullan@cs.ucl.ac.uk

Department of Computer Science,
Centre for Medical Image Computing,
Centre for Inverse Problems
University College London

Lecture 15

Convex constraint optimization problem

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

where

- $f : \mathcal{D} \to \mathbb{R}$ is convex, twice continuously differentiable function, $\mathcal{D} \subset \mathbb{R}^n$ is convex
- $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots m$ are convex, twice continuously differentiable functions
- $A \in \mathbb{R}^{p \times n}$ with $\operatorname{rank} A = p < n$.

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 optimal $\lambda^\star \in \mathbb{R}^m$, $\nu^\star \in \mathbb{R}^p$, which together with $x^\star$ satisfy the KKT conditions

$$
\begin{aligned}
\nabla f(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(x^\star) + A^{\mathrm{T}} \nu^\star &= 0, && \text{(KKT)} \\
Ax^\star &= b, \\
f_i(x^\star) &\le 0, \quad i = 1, \dots, m, \\
\lambda^\star &\ge 0, \\
\lambda_i^\star f_i(x^\star) &= 0, \quad i = 1, \dots. m.
\end{aligned}
$$

Interior point methods

Interior point methods solve either

- the problem (COP) by applying Newton's method to a sequence of equality constraint problems.

- the conditions (KKT) by applying Newton's method to a sequence of modified versions of the KKT conditions.

We will consider the *barrier method* and the *primal-dual interior-point method*.

Logarithmic barrier function

Rewrite (COP) making the inequality constraints implicit

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

$$\text{subject to} \quad Ax = b,$$

where $I_- : \mathbb{R} \to \mathbb{R}$ is the indication function for the nonpositive reals

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

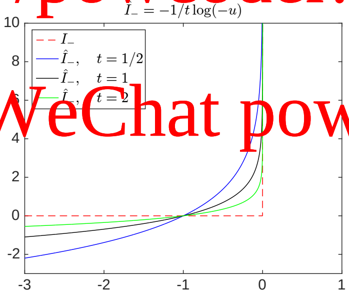
I_- is non-differentiable thus we need a smooth approximation before Newton method can be applied.

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 $u > 0$.
- Unlike I_-, $\hat{I}_-$ is differentiable and closed i.e. it increases to ∞ as u increases to 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.

Logarithmic barrier

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

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

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

$$\text{subject to} \quad Ax = b.$$

We assume that (CENT) has a unique solution for each $t > 0$, and denote this solution with $x^\star(t)$.

The set of points $x^\star(t)$, $t > 0$ is called the **central path**. The points on central path are characterised by the following necessary and sufficient *centrality conditions*:

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

$$Ax^\star(t) = b, \qquad f_i(x^\star(t)) < 0, \; i = 1, \dots, m,$$

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

$$\begin{aligned} 0 &= t\nabla f(x^\star(t)) + \nabla\phi(x^\star(t)) + A^{\mathrm{T}}\hat{\nu} \qquad \text{(CENT:COND)} \\ &= t\nabla f(x^\star(t)) + \sum_{i=1}^{m} \frac{1}{-f_i(x^\star(t))} \nabla f(x^\star(t)) + A^{\mathrm{T}}\hat{\nu} \end{aligned}$$

Example: LP with inequality constraints

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

The logarithmic barrier:

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

The gradient and Hessian:

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

Since x is strictly feasible, we have $b_i - a_i^{\mathrm{T}} x > 0$ and the Hessian is nonsingular iff A has rank n.

The centrality condition (CENT:COND): $(\nabla \phi(x^\star(t)) \parallel -c)$

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

Example: central path

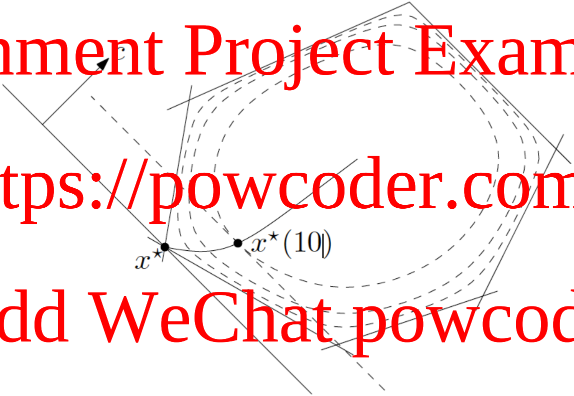


Figure: Boyd Vandenberghe Fig. 11.2

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

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

is dual feasible.

Proof: $\lambda^\star(t) > 0$ because $x^\star(t)$ is strictly feasible $f_i(x^\star(t)) < 0$ and from optimality conditions (CENT:COND) we read out that $x^\star(t)$ 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

$$
\begin{aligned}
g(\lambda^\star(t), \nu^\star(t)) &= f(x^\star(t)) + \sum_{i=1}^{m} \underbrace{\lambda_i^\star(t)}_{=-\frac{1}{t f_i(x^\star(t))}} f_i(x^\star(t)) \\
&\quad + \nu^\star(t)^{\mathrm{T}} \underbrace{(Ax^\star(t) - b)}_{=0} \\
&= f(x^\star(t)) - m/t \quad \to \text{duality gap}
\end{aligned}
$$

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

$$f(x^\star(t)) - p^\star \leq m/t$$

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

Interpretation via KKT conditions

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

$$
\begin{aligned}
\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + A^{\mathrm{T}} \nu &= 0, \qquad \text{(KKT:CENT)}\\
Ax &= b,\\
f_i(x) &\leq 0, \quad i = 1, \dots, m,\\
\lambda &\geq 0,\\
-\lambda_i f_i(x) &= 1/t, \quad i = 1, \dots, m.
\end{aligned}
$$

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) (linear equality constraint problem) reads

$$\begin{bmatrix} t\nabla^2 f(x) + \nabla^2\phi(x) & A^{\mathrm{T}} \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_n \\ \nu_n \end{bmatrix} = - \begin{bmatrix} t\nabla f(x) + \nabla\phi(x) \\ 0 \end{bmatrix}.$$

Here we assumed feasibility.

We can interpret these Newton step for (CENT) as Newton for directly solving the modified (KKT:CENT) in a particular way.

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

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

To find the Newton step for the solution of the nonlinear equations above, we form the Taylor expansion for the nonlinear term

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

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

$$tH\Delta x_n + A^{\mathrm{T}}\nu_n = -tg, \quad A\Delta x_n = 0.$$

Comparing to the Newton step for (KKT:CENT) yields

$$v = \Delta x_n, \quad \nu = (1/t)\nu_n.$$

This shows that the Newton step for the centring problem (CENT) 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

  Centering step:
  obtain $x^\star(t)$ by minimising (CENT) starting from x^s

  Update $x^s = x^\star(t)$

  if $m/t < \epsilon$ **then**

    break {stopping criterium ϵ-sub optimal point}

  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 centering** is not necessary since the central path has no significance beyond that it leads to the solution of the original problem (COP) as $t \to \infty$.
  - **Inexact centering** will still produce a convergent sequence, however the $\lambda^\star(t), \nu^\star(t)$ are not exactly dual feasible (can be corrected).
  - The difference in cost between the exact and good approximate centering is marginal (few Newton steps), so centering is usually assumed exact.

- **Choice of μ:** trade of between the number of outer (centering) and inner (Newton) iterations.
  - **Small $\mu \approx 1$:** good initial guess for Newton i.e. small number of inner iterations closely following the central path but a large number of outer iterations to reach desire accuracy ϵ.
  - **Larger μ:** only a few outer iterations but with a large number of inner iterations straying from the central path.
  - In practice for a large range of $\mu \in (3, 100)$ these effects balance each other yielding approximately same total number of Newton iterations. Values around 10-20 seem to work well. For best worse-case bound on total Newton steps set μ close to 1.

- **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 set $t^{(0)} = m/\eta$ (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

$$\inf_{t,\nu} \|t\nabla f(x^{(0)}) + \nabla\phi(x^{(0)}) + A^{\mathrm{T}}\nu\|,$$

    a measure of deviation of $x^{(0)}$ from $x^\star(t)$ (least squares problem for t, ν).

- **Infeasible Newton method** for centering step. Choose $x^{(0)} \in \mathcal{D}$, $f_i(x^{(0)}) < 0, i = 1, \dots, m$ but not necessarily $Ax^{(0)} = b$. 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 find that the constraints are infeasible).

Consider the set of inequalities and equalities

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

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 $x^{(0)}$.

Phase I: max

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

$$\begin{aligned} \min \quad & s \\ \text{subject to} \quad & f_i(x) \le s, \quad i = 1, \dots, m \\ & Ax = b \end{aligned} \tag{PH1:MAX}$$

s: bound on the the maximum infeasibility of the inequalities. The goal is to drive this maximum below 0.

The problem (PH1:MAX) is always strictly feasible. Thus we can initialise with $x = x^{(0)}$ and for s with any number larger than $\max_{i=1,\dots,m} f_i(x^{(0)})$ 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 $s < 0$.
- $p_I^\star > 0$: (FEAS) are infeasible.
  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 $p_I^\star > 0$.
- If $p_I^\star = 0$ and the minimum is attained at $x^\star$ and $s^\star = 0$, then the set of inequalities is feasible, but not strictly feasible. If $p^\star = 0$ and the minimum is not attained, the inequalities are infeasible.

Phase I: sum

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

- For a fixed x, the optimal value of s_i is $\max\{f_i(x), 0\}$. Thus we are minimising a sum of infeasibilities.
- The optimal value is 0 and achieved iff the original set of equalities and inequalities is feasible.
- When the system of equalities and inequalities is infeasible, often the solution violates only a small number of constraints 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 $x^{(0)} \in \mathcal{D} \cap \prod_{i=1}^m \mathrm{dom} f_i$ with $Ax^{(0)} = b$.

Modified phase I optimisation problem

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

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

Central path for this modified problem $(x^\star(\bar{t}), s^\star(\bar{t})), \bar{t} > 0$

$$\sum_{i=1}^m \frac{1}{s - f_i(x)} = \bar{t}, \quad \frac{1}{M - f(x)} \nabla f(x) + \sum_{i=1}^m \frac{1}{s - f_i(x)} \nabla f_i(x) + A^{\mathrm{T}} \nu = 0.$$

If (x, s) with $s = 0$ is on this central path, it is also on the central path for (COP) if the latter is strictly feasible ($s = 0$)

$$t \nabla f(x) + \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x) + A^{\mathrm{T}} \nu = 0$$

with $t = 1/(M - f(x))$.

We expresse (COP) in equivalent form

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

Start the barrier method using infeasible Newton method so solve

$$
\begin{aligned}
\min \quad & tf(x) - \sum_{i=1}^{m} \log(s - f_i(x)) \\
\text{subject to} \quad & Ax = b, \quad s = 0,
\end{aligned}
$$

which can be initialised with any $x \in \mathcal{D}$ and any $s > \underbrace{\max_i f_i(x)}_{\textit{infeasibility}}$.

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

Finding a point in the domain $\mathcal{D}$

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

$$\min \quad tf(x + z_0) - \sum_{i=1}^{m} \log(s - f_i(x + z_i))$$

$$\text{subject to} \quad Ax = b, \; s = 0, \; z_0 = 0, z_1 = 0, \ldots, z_m = 0,$$

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.

Characteristic performance

- 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 produce a certificate of infeasibility grows.
- When the problem is exactly on the boundary between strictly feasible and infeasible, for example, feasible but not strictly feasible, the cost becomes infinite.
- Typically the infeasible start Newton method works very well provided the inequalities are feasible, and not very close to the boundary between feasible and infeasible.
- When the feasible set is just barely nonempty, a phase I method is far better choice. Phase I method gracefully handles the infeasible case; the infeasible start Newton method, in contrast, simply fails to converge.

Primal-dual interior point method

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. At each iteration, both the primal and dual variables are updated.
- The search directions in a primal-dual interior-point method are obtained from Newton's method, applied to modified KKT equations (i.e., the optimality conditions for the logarithmic barrier centering problem). The primal-dual search directions are similar to, but not quite the same as, the search directions that arise in the barrier method.
- In a primal-dual interior-point method, the primal and dual iterates are not necessarily feasible.
- Usually more efficient than barrier methods, and do not require strict feasibility.

Primal-dual search direction

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

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

with $t > 0$, $F : \mathbb{R}^n \to \mathbb{R}^m$ and its Jacobian

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

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

$$r_t(y + \Delta y) \approx r_t(y) + Dr_t(y)\Delta y = 0,$$

where $\Delta y = (\Delta x, \Delta\lambda, \Delta\nu)$ is the *primal-dual search direction*.

Written in terms of x, λ, ν:

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

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

From the second block

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

and substitute into the first block

$$\begin{bmatrix} H_{\mathrm{pd}} & A^{\mathrm{T}} \\ A & 0 \end{bmatrix}\begin{bmatrix} \Delta x_{\mathrm{pd}} \\ \Delta\nu_{\mathrm{pd}} \end{bmatrix} = -\begin{bmatrix} r_{\mathrm{dual}} + J(x)^{\mathrm{T}}\operatorname{diag}(f(x))^{-1} r_{\mathrm{cent}} \\ r_{\mathrm{pri}} \end{bmatrix}$$

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

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)

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

where

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

Multiplying first block by $1/t$ and changing variables $\Delta\nu_{\text{bar}} = (1/t)\nu_{\text{bar}} - \nu$

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

The right hand sides are identical.

The only difference are

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

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

When x, λ, ν satisfy $-f_i(x)\lambda_i = 1/t$, the coefficient matrices (and 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 kth step, as we do in the outer loop of the barrier method.
- Instead, we define the **surrogate duality gap**, for any x that satisfied $F(x) < 0$ and $\lambda \geq 0$ as

$$\eta(x, \lambda) = -F(x)^{\mathrm{T}} \lambda$$

- 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: $\mu > 1$
Require: Tolerances $\epsilon_{\text{feas}} > 0, \epsilon > 0$

repeat
  Determine t: $t := \mu m / \eta$
  Compute primal-dual search direction Δy_{pd}
  Line search: determine step length $s > 0$ and set $y := y + s \Delta y_{\text{pd}}$
until $\|r_{\text{prim}}\| \le \epsilon_{\text{feas}}$, $\|r_{\text{dual}}\| \le \epsilon_{\text{feas}}$ and $\eta \le \epsilon$

Remarks

- 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 in the barrier method).

- Values of the parameter μ on the order of 10 appear to work well.

- The primal-dual interior-point algorithm terminates when x is primal feasible and λ, ν are dual feasible (within the tolerance 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

$$\|r_t(x+s\Delta x_{\mathrm{pd}}, \lambda+s\Delta\lambda_{\mathrm{pd}}, x+s\Delta\nu_{\mathrm{pd}})\| \leq (1-\alpha s)\|r_t(x,\lambda,\nu)\|.$$

  Common choices for backtracking parameters are same as for Newton method α in the range 0.01 to 0.1 and ρ 0.3 to 0.8.
- One iteration of the primal-dual interior-point algorithm is the same as one step of the infeasible Newton method, applied to solving $r_t(x,\lambda,\nu) = 0$, but modified to ensure $\lambda > 0$ and $F(x) < 0$ (or, equivalently, with $\operatorname{dom} r_t$ 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.