Interior Point implementation note

June 25, 2023

General form:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f_0\left(x\right) \\ \text{s.t.} & f_i\left(x\right) \leq 0 \\ & Ax = b \end{array}$$

Centering Problem with log-barrier:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f_0\left(x\right) - \frac{1}{t}\sum_{i=1}^{m} \log\left(-f_i\left(x\right)\right) \\ \text{s.t.} & Ax = b \end{array}$$

Our case:

$$\begin{array}{ll} \underset{z}{\text{minimize}} & \frac{1}{2} z^T H z \\ s.t. & A_i z \leq b_i \\ & A_e z = b_e \end{array}$$

With $z = \left(x_0 - x_0^{\text{ref}}, u_0, \ldots, x_{N-1} - x_{N-1}^{\text{ref}}, u_{N-1}, x_N - x_N^{\text{ref}}\right)$ and

$$H = \begin{bmatrix} W & 0 & 0 & 0 \\ & \ddots & & \\ 0 & & 0 & 0 \\ 0 & 0 & W & 0 \\ 0 & 0 & 0 & Q_N \end{bmatrix}, W = \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix}, A_e = \begin{bmatrix} I & 0 & 0 & \cdots & \cdots & 0 \\ -A & -B & I & 0 & \cdots & 0 \\ & & & & \ddots & \vdots \\ 0 & -A & -B & I & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & -A & -B & I \end{bmatrix}, b_e = \begin{bmatrix} x_{\text{init}} - x_0^{\text{ref}} \\ A x_0^{\text{ref}} - x_1^{\text{ref}} \\ A x_1^{\text{ref}} - x_2^{\text{ref}} \\ \vdots \\ A x_{N-1}^{\text{ref}} - x_N^{\text{ref}} \end{bmatrix}$$

$$\tilde{A}_i = \begin{bmatrix} C & & & & & \\ & I_m & & & & \\ & & \ddots & & & \\ & & & C & & \\ & & & & I_m & \\ & & & & & C \end{bmatrix}, b_{i_1} = \begin{bmatrix} y_{\max} - C x_0^{\text{ref}} \\ u_{\max} \\ \vdots \\ y_{\max} - C x_{N-1}^{\text{ref}} \\ u_{\max} \\ y_{\max} - C x_N^{\text{ref}} \end{bmatrix}, b_{i_2} = \begin{bmatrix} -y_{\min} + C x_0^{\text{ref}} \\ -u_{\min} \\ \vdots \\ -y_{\min} + C x_{N-1}^{\text{ref}} \\ -u_{\min} \\ -y_{\min} + C x_N^{\text{ref}} \end{bmatrix}, A_i = \begin{bmatrix} \tilde{A}_i \\ -\tilde{A}_i \end{bmatrix}, b_i = \begin{bmatrix} b_{i_1} \\ b_{i_2} \end{bmatrix}$$

1 Infeasible start Newton Method for MPC centering problem

- **We apply a log-barrier penalty:**

$$\begin{array}{ll} \underset{z}{\text{minimize}} & \frac{1}{2}z^T Hz - \frac{1}{t}\sum\limits_{i=1}^{m}\log\left(b_i - a_i^T z\right) \\ \text{s.t.} & A_e z = b_e \end{array}$$

- a_i^T is the i^{th} row of A_i
- $t > 0$ is a parameter that sets the accuracy of the approximation. As t increases the approximation becomes more accurate

- **Centering Path problem:**

$$\begin{array}{ll} \underset{z}{\text{minimize}} & \frac{t}{2}z^T Hz - \sum\limits_{i=1}^{m}\log\left(b_i - a_i^T z\right) \\ \text{s.t.} & A_e z = b_e \end{array}$$

We note $f(z) = f_0(z) + \phi(z)$

With $f_0(z) = \frac{t}{2}z^T Hz, \phi(z) = -\sum\limits_{i=1}^{m}\log\left(b_i - a_i^T z\right)$,

- $\nabla f_0(z) = tHz$
- $\nabla^2 f_0(z) = tH$
- $\nabla\phi(z) = \sum\limits_{i=1}^{m}\frac{1}{b_i - a_i^T z}a_i = A^T d$ with $d = \left(\frac{1}{b_1 - a_1^T z}, \dots, \frac{1}{b_m - a_m^T z}\right)$
- $\nabla^2\phi(z) = \sum\limits_{i=1}^{m}\frac{a_i^T}{\left(b_i - a_i^T z\right)^2}a_i = A_i^T \operatorname{diag}(d)^2 A_i$

We summarize as:

$$\boxed{\begin{array}{c} \nabla f(z) = tHz + A_i^T d \\ H_f = \nabla^2 f(z) = tH + A_i^T \operatorname{diag}(d) A_i \\ \text{with } d = \left(\frac{1}{b_1 - a_1^T z}, \dots, \frac{1}{b_m - a_m^T z}\right) \end{array}}$$

- We note that $b_i \succ 0$ when the problem is setup correctly
  - This means $z = 0 \in \operatorname{dom} f$ and can be used for starting infeasible start newton method

- **We use the primal-dual method computing residual:**
  - Lagrangian is: $L(z, \nu) = f(z) + \nu^\top (A_e z - b_e)$
  - We want to drive the residuals $(\nabla_z L, \nabla_\nu L)$ to zero
  - We note $y = (z, \nu), r(y) = \left(\nabla f(z) + A_e^\top \nu, A_e z - b_e\right)$
  - We linearize $r(y + \Delta y) = r(y) + Dr(y)\Delta y = 0$
  - $Dr(y) = \begin{bmatrix} \nabla^2 f & A_e^\top \\ A_e & 0 \end{bmatrix}$

– So we get the KKY system

$$\begin{bmatrix} \nabla^2 f & A_e^\top \\ A_e & 0 \end{bmatrix} \begin{bmatrix} \Delta z \\ \Delta \nu \end{bmatrix} = - \begin{bmatrix} \nabla f(z) + A_e^\top \nu \\ A_e z - b_e \end{bmatrix}$$

– $H_f = \nabla^2 f$ is a diagonal matrix, easy to invert

– **A_e and H_f are sparse**

– So we can just use a sparse linear solver \ with SparseArrays.jl

- **We could solve the KKT system by elimination**. H is a diagonal matrix

$$\begin{cases} H_f \Delta z + A_e^\top \Delta \nu = -g \\ A_e \Delta z = -h \end{cases}$$

$$\begin{cases} \Delta z = -H_f^{-1}\left(g + A_e^\top \Delta \nu\right) \\ A_e \Delta z = -h \end{cases}$$

$$\begin{cases} \Delta z = -H_f^{-1}\left(g + A_e^\top \Delta \nu\right) \\ -A_e H_f^{-1}\left(g + A_e^\top \Delta \nu\right) = -h \end{cases}$$

$$\begin{cases} \Delta z = -H_f^{-1}\left(g + A_e^\top \Delta \nu\right) \\ -A_e H_f^{-1} A_e^\top \Delta \nu = -h + A_e H_f^{-1} g \end{cases}$$

$$\boxed{\begin{cases} \left(A_e H_f^{-1} A_e^\top\right) \Delta \nu = h - A_e H_f^{-1} g \\ H_f \Delta z = -\left(g + A_e^\top \Delta \nu\right) \end{cases}}$$

– As $H_f \succ 0$, we have $A_e H_f^{-1} A_e^\top \succ 0$ Positive Definite (of size $? \times ?$)

– We can compute by **Cholesky factorization** $AH^{-1}A^\top = LL^\top$ in $\frac{1}{3}?^3$ flops

– Thus solving the KKT system is $\mathcal{O}\left(?^3\right)$ to provide the primal and dual Newton steps $\Delta x, \Delta \nu$

- **Backtracking Line search:**

– We used to have for a function f , reducing f enough along a search direction

  * Stopping Condition: $f(x + t\Delta x) \leq f(x) + \alpha \nabla f(x)^\top (t\Delta x)$
  * Note that $\nabla f(x)^\top (t\Delta x) \leq 0$ in a decrease direction
  * In words: we accept a decrease in f of α percent of the prediction in the linear extrapolation

– For the primal-dual method, the norm of the residual decreases in the Newton direction:

* $\frac{d}{dt} \left\| r\left(y + t\Delta y_{pd}\right) \right\|_2^2 \Big|_{t=0} = 2r\left(y\right)^\top Dr\left(y\right)\Delta y_{pd} = -2r\left(y\right)^\top r\left(y\right)$
  as $Dr\left(y\right)\Delta y = -r\left(y\right)$
* $\frac{d}{dt} \left\| r\left(y + t\Delta y_{pd}\right) \right\|_2^2 \Big|_{t=0} = -2\left\| r\left(y\right) \right\|_2^2$
* As for $h = g \circ f$ we have $Dh = Dg\left(f\right)Df$ we get
  · With $g = \sqrt{\cdot}$ and $f = \left\| r\left(y + t\Delta y_{pd}\right) \right\|_2^2$
  · $\frac{d}{dt} \left\| r\left(y + t\Delta y_{pd}\right) \right\|_2 \Big|_{t=0} = \frac{d}{dt}\sqrt{\left\| r\left(y + t\Delta y_{pd}\right) \right\|_2^2}\Big|_{t=0}$
  · $\frac{d}{dt} \left\| r\left(y + t\Delta y_{pd}\right) \right\|_2 \Big|_{t=0} = \frac{1}{2\sqrt{\left\| r(y) \right\|_2^2}}\left(-2\left\| r\left(y\right) \right\|_2^2\right) = -\left\| r\left(y\right) \right\|_2$

$$\boxed{\frac{d}{dt} \left\| r\left(y + t\Delta y_{pd}\right) \right\|_2 \Big|_{t=0} = -\left\| r\left(y\right) \right\|_2}$$

– The Stopping Condition for our merit function is:

$$\boxed{\left\| r\left(y + t\Delta y_{pd}\right) \right\|_2 \leq \left\| r\left(y\right) \right\|_2 - \alpha t \left\| r\left(y\right) \right\|_2}$$

Algorithm 10.2 *Infeasible start Newton method.*

given starting point $x \in \mathbf{dom}\, f$, ν, tolerance $\epsilon > 0$, $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$.
repeat
1. Compute primal and dual Newton steps Δx_{nt}, $\Delta \nu_{\mathrm{nt}}$.
2. *Backtracking line search on* $\|r\|_2$.
   $t := 1$.
   while $\|r(x + t\Delta x_{\mathrm{nt}}, \nu + t\Delta\nu_{\mathrm{nt}})\|_2 > (1 - \alpha t)\|r(x, \nu)\|_2$, $\quad t := \beta t$.
3. *Update.* $x := x + t\Delta x_{\mathrm{nt}}$, $\nu := \nu + t\Delta\nu_{\mathrm{nt}}$.

until $Ax = b$ and $\|r(x, \nu)\|_2 \leq \epsilon$.

---

2 Primal-dual method

$$\begin{array}{ll} \underset{z}{\text{minimize}} & \frac{1}{2}z^T H z \\ s.t. & A_i z \leq b_i \\ & A_e z = b_e \end{array}$$

$L\left(z, \lambda, \nu\right) = \frac{1}{2}z^T H z + \lambda^T\left(A_i z - b_i\right) + \nu^T\left(A_e z - b_e\right)$
$\nabla L_z\left(z, \lambda, \nu\right) = Hz + A_i^T\lambda + A_e^T\nu$

Perturbated KKT conditions:

$$
\begin{cases}
Hz + A_i^T \lambda + A_e^T \nu = 0 \\
-\lambda^T (A_i z - b_i) = \frac{1}{t}\mathbf{1} & \text{complementary slackness } = 0 \\
A_i z - b_i \leq 0 \\
\lambda \geq 0
\end{cases}
$$

We can express:

$$
r_t(z, \lambda, \nu) = \begin{bmatrix} Hz + A_i^T \lambda + A_e^T \nu \\ -\lambda^T (A_i z - b_i) - \frac{1}{t}\mathbf{1} \\ A_e z - b_e \end{bmatrix}
$$

If z, λ, ν satisfy $r_t(z, \lambda, \nu) = 0$ and $A_i z - b_i \leq 0$, then $x = x^*(t), \lambda = \lambda^*(t)$, and $\nu = \nu^*(t)$.

In particular, x is primal feasible, and λ, ν are dual feasibke with duality gap $\frac{m}{t}$

We note:

- $r_{\text{dual}} = Hz + A_i^T \lambda + A_e^T \nu$ the dual residual
- $r_{\text{pri}} = A_e z - b_e$ the primal residual
- $r_{\text{cent}} = \lambda^T (A_i z - b_i) - \frac{1}{t}\mathbf{1}$ the centrality residual, i.e., the residual for the modified complementary conditions.

We consider the Newton step for solving the nonlinear equations $r_t(z, \lambda, \nu) = 0$ for fixed t at a point (z, λ, ν) that satisfies $A_i z \preceq b_i, \lambda \succeq 0$

We will denote the current point and Newton step as

$$
y = (z, \lambda, \nu), \Delta y = (\Delta z, \Delta \lambda, \Delta \nu)
$$

The Newton step is characterized by the linear equations

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

Leading to

$$
\Delta y = -Dr_t(y)^{-1} r_t(y)
$$

Which is

$$
\begin{bmatrix} H & A_i^T & A_e^T \\ -A_i^T \lambda & -(A_i z - b_i) & 0 \\ A_e & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta z \\ \Delta \lambda \\ \Delta \nu \end{bmatrix} = - \begin{bmatrix} r_{\text{dual}} \\ r_{\text{cent}} \\ r_{\text{pri}} \end{bmatrix}
$$