Interior Point methods

problem:

$$\min \; f$$
$$\text{st} \; g_i(x) \leq 0$$

(assume there is some x st $g_i(x) < 0, \forall i$)

strategy: we wish to avoid the boundary because progress on the boundary may be slow. It's hard to take large steps, because the active constraints might change. If we overstep to produce an infeasible pt., projecting back could be tough.

Introduce a penalty barrier

$$\phi(x) = -\sum_{i=1}^{m} \cancel{\left(\frac{1}{t}\right)} \log\left(-\underset{g}{\cancel{f}}_i(x)\right)$$

As long as $g_i(x) < 0$, this is finite. (2)

at $g_i(x) = 0$, $+ve\ \infty$

for $g_i(x)$ very negative (i.e. far from boundary) this is small.

now $\min \; f(x) + \frac{1}{t} \cdot \varphi(x)$

(multiply through by t)

<u>So</u>:

$$\min \; t\, f(x) + \varphi(x).$$

write $x^*(t) = \underset{x}{\arg\min} \; t f(x) + \varphi(x)$

This is called the <u>central path</u>

large t : $x^*(t)$ way in interior

small t : could be at boundary

Notice

1) f convex, g_i convex means

$tf + \varphi$ is convex.

2) usual to have linear ~~in~~equalities as well

$$x^*(t) = \operatorname{argmin} \; tf + \varphi$$

$$\text{s.t.} \quad Ax = b$$

3) at $x^*(t)$, we must have

- $Ax^* = b$ (because this is a min).
- $g_i(x^*) < 0$
- $t \nabla f(x^*) + \nabla \varphi(x^*) + \nu^T A = 0$

4

which gives

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

4) there are λ^*, ν^* ← dual feasible corresponding to x^*

write $\lambda_i^* = \dfrac{-1}{t\, g_i(x^*)}$

this is +ve.

~~rewrite optimality as.~~

write $\nu^* = \dfrac{\nu}{t}$

then rewrite optimality as

$$\nabla f + \sum_i \lambda_i^* \nabla g_i(x^*) + \nu^{*T} A = 0$$

but this is condition for a Lagrangian —
so x^* ~~is~~ ~~is~~ minimizes

$$f + \sum_i \lambda_i^* g_i + \nu^{*T}(Ax - b)$$

for λ^*, ν^*

$\Rightarrow \lambda^*, \nu^*$ is dual-feasible.

This is important, because we have that the dual $g(\lambda, \nu)$ is finite at λ^*, ν^*

We can evaluate the dual at λ^*, ν^*

$$g(\lambda^*, \nu^*) = \inf_x \mathcal{L}(x, \lambda^*, \nu^*)$$

$$= \mathcal{L}(x^*, \lambda^*, \nu^*)$$

$$= f + \underbrace{\sum_i \left[\frac{-1}{t\, g_i(x^*)}\right] g_i(x^*)}_{-m/t} + \nu^{*T} \underbrace{[Ax^* - b]}_{0}$$

$$= f(x^*) - m/t$$

This is a lower bound on soln

$f(x^*)$ is upper bound

So its trapped in a range $[f(x^*) - m/t, f(x^*)]$

we have

$$\min \; t f(x) + \phi(x)$$

$$\text{st} \quad Ax = b$$

<u>KKT</u>:

$$Ax = b$$

$$g_i(x) \leq 0$$

$$\lambda_i \geq 0$$

$$\nabla f + \sum_i \lambda_i \nabla g_i + \nu^T A = 0$$

$$\underbrace{-\lambda_i g_i \cancel{f_i}(x) = 1/t}_{\uparrow}$$

from defn of λ^*

Notice that these look like

KKT for original problem, with a change in the complementarity cond.

Alg: start with feasible x, $t^0 > 0$, $\mu > 1$. $\varepsilon > 0$

- $x^* = \arg\min\limits_{st} \; tf + \varphi$ $\quad Ax = b$

(ALM?)

- $x = x^*$
- if $\frac{m}{t} < \varepsilon$ quit else $t = \mu t$.

What about SDP?

consider
$$\phi(X) = -\log[\det(X)]$$

1) X has all large eigenvalues $\rightarrow$ ϕ is positive small

2) $\det X \rightarrow 0^{+}$ $\Rightarrow$ $\phi(x) \rightarrow +\infty$

gradients:
$$\frac{\partial}{\partial x_{ij}}\left(\log \det X\right) = \frac{1}{\det(X)}\left[\text{coeff of } x_{ij} \text{ in det}\right]$$

Sometimes written.
$$\frac{\partial}{\partial X}(\log \det X) = X^{-1}$$

Primal-dual methods

recall that a path following method solved KKT

$$A x = b$$
$$g_i(x) \le 0$$
$$\lambda_i \geq 0$$
$$\nabla f + \sum_i \lambda_i \nabla g_i + \nu^T A = 0$$
$$-\lambda_i f_i(x) = 1/t$$

But we solved for x^* for given t,
then recovered λ^*, ν^*.
— we could instead view KKT
as a system of equs in λ, ν, x
for given t, solve, make t
~~smaller~~, etc.
bigger

how to solve?

a) they're nonlinear, even if inequality constraints are linear

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

Nasty complementarity cond.

b). see this as root finding → Newton.

find $G(u^*) = 0$

$u^{(n)}$: then

$$G(u^{(n)} + \Delta u) \approx G(u^{(n)}) + J_G \cdot \Delta u$$

so $J_G \Delta u = -G$.

Our system

$$\nabla f + \lambda^T J_g + \nu^T A$$ ← dual residual

If this is 0, λ, ν are df

$$-\operatorname{diag}(\lambda) \cdot g - \frac{1}{t} \cdot 1$$ ← centering residual

$$Ax - b$$ ← primal residual

- centering residual expresses whether point is on central path
- primal residual — is x feasible

r_d

r_c

r_p

Newton's method gives:

$$\begin{bmatrix} H_f + \sum_i \lambda_i H_{g_i} & J_g & A^T \\ -\mathrm{diag}(\lambda) J_g & -\mathrm{diag}(g) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta \nu \end{bmatrix} = \begin{bmatrix} -r_d \\ -r_c \\ -r_p \end{bmatrix}$$

But how good is the point?

we don't necessarily have either primal feasible x ~~or~~ dual feasible λ, ν — so we can't get duality gap. Instead, work with surrogate

duality gap:

- assume x is primal feasible — then primal has value $f(x)$.
- assume λ, ν are dual feasible, ~~and we're at~~ dual has value

$$f(x) + \lambda^T g$$

so gap is primal − dual

$$\eta = -\lambda^T g$$

(Notice here we're assuming $g \leq 0$, $\lambda \geq 0$)

notice also that this is related to complementarity

$$-\lambda_i g_i = \frac{1}{t}$$

so $\hat{\eta} = \frac{m}{t}$ OR $t = \frac{m}{\hat{\eta}}$

Algorithm:

Start with feasible x $(g(x) < 0)$

" $\lambda > 0$

$\mu > 1$

$\varepsilon_f > 0$

$\varepsilon > 0$

Iterate

- $t = \mu \dfrac{m}{\hat{\eta}}$
- get line search dir by ~~Newton~~ linear alg on system above
- line search; find $s > 0$

$$\begin{matrix} x \\ \lambda \\ \nu \end{matrix} \rightarrow \begin{matrix} x \\ \lambda \\ \nu \end{matrix} + s \begin{matrix} \Delta x \\ \Delta \lambda \\ \Delta \nu \end{matrix}$$

- until $\|r_f\|_2 \leq \varepsilon_{feas}$, $\|r_d\| \leq \varepsilon_{feas}$, $\hat{\eta} < \varepsilon$.

line search:

- look for $s\begin{pmatrix}\Delta x\\ \Delta\lambda\\ \Delta\nu\end{pmatrix}$ that gets smallest value of norm of residual
- ensure $\lambda > 0$, $g(x) < 0$

strategy

- find largest $s = s^*$ that gives $\lambda > 0$
- now search back by

$$s^{(n+1)} = \beta s^{(n)} \qquad 0 < \beta < 1$$

looking for sufficient improvement in residual