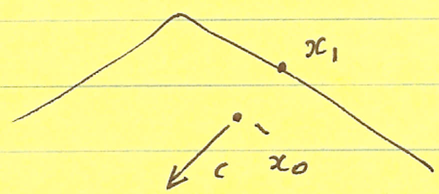
Interior point methods:

- Why is it worth knowing L.M.'s?
  - because they make it possible to follow the boundary.

eg. LP

$$\min \ c'x \quad \text{st} \quad Ax = b, \quad x \geq 0$$



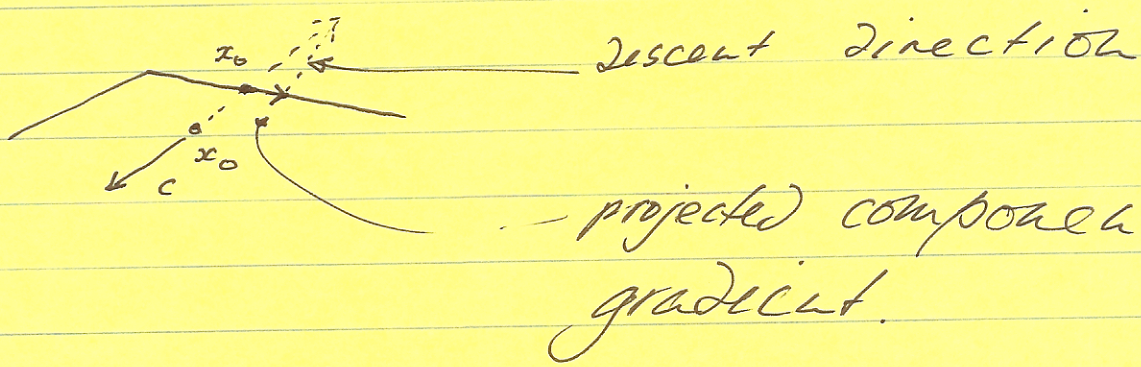
- x_1 is obvious, but what next?
- we must move <u>along</u> boundary.

direction of motion

= Gradient − $\underbrace{\text{normal components}}_{\text{and these are given by L.M.'s}}$

However it is not a good idea to move <u>directly</u> to the boundary.

- Why? steps may then be small.



- General idea of interior point methods

  - Bias step to stay in interior for long steps

Continuation view:

$\min f$

st $\quad c_E = 0 \qquad c_I - s = 0 \qquad s \geq 0$

This gives

$$\nabla f - J_E' y - J_I' z = 0$$

$$Sz - \mu 1 = 0 \qquad (S \text{ is } \operatorname{diag}(s))$$

$$c_E = 0 \qquad c_I - s = 0$$

$$s \geq 0 \quad , \quad z \geq 0$$

and $\mu = 0$ for exact system.

Penalty Barrier view:

$$\min \quad f(x) - \mu \sum_i \log s_i$$

$$\text{st} \quad c_E = 0 \qquad c_I - s = 0$$

KKT

$$\nabla_x f - J_E' y - J_I' z = 0$$

$$-\mu S^{-1} 1 + z = 0$$

$$c_E = 0 \qquad c_I - s = 0$$

Notice S^{-1} gets nasty for $s_i \to 0$

$\therefore$ multiply 2nd eqn by S

to get $Sz - \mu 1 = 0$

as previous

We can now apply Newton's method.

$$\begin{bmatrix} H_x\mathcal{L} & 0 & -J_E' & -J_I' \\ 0 & Z & 0 & S \\ J_E & 0 & 0 & 0 \\ J_I & -I & 0 & 0 \end{bmatrix} \begin{matrix} \Delta x \\ \Delta s \\ \Delta y \\ \Delta z \end{matrix} = \begin{bmatrix} \nabla f - J_E' y - J_I' z \\ Sz - \mu 1 \\ C_E \\ C_I - s \end{bmatrix}$$

where H_x is hessian of Lagrangian.

- get step

$$\begin{pmatrix} x \\ s \end{pmatrix} \rightarrow \begin{pmatrix} x \\ s \end{pmatrix} + \alpha_s \begin{pmatrix} \Delta x \\ \Delta s \end{pmatrix}$$

$$\begin{pmatrix} y \\ z \end{pmatrix} \qquad \begin{pmatrix} y \\ z \end{pmatrix} + \alpha_z \begin{pmatrix} \Delta y \\ \Delta z \end{pmatrix}$$

so neither set of vars goes so boundary too fast.

where $\alpha_s = \max\{\alpha \in (0,1) : s + \alpha \Delta s \geq (1-\tau)s\}$

$\alpha_z = \max\{\alpha \in (0,1) : z + \alpha \Delta z \geq (1-\tau)z\}$

Linear algebra:

rewrite in symmetric form, with $\Sigma = S^{-1} Z$

$$
\begin{bmatrix} H_x L & 0 & J_E' & J_I' \\ 0 & \Sigma & 0 & -I \\ J_E & 0 & 0 & 0 \\ J_I & -I & 0 & 0 \end{bmatrix}
\begin{bmatrix} \Delta x \\ \Delta s \\ -\Delta y \\ -\Delta z \end{bmatrix}
=
\begin{bmatrix} \nabla f - J_E' y - J_I' z \\ z - \mu S^{-1} 1 \\ c_E \\ c_I - s \end{bmatrix}
$$

notice we can eliminate Δs easily.

$$
\begin{aligned}
\Delta s &= \Sigma^{-1}\left[z - \mu S^{-1} 1 - \Delta z \right] \\
&= \underbrace{Z^{-1} S z}_{\to\ \text{diag!}\ s} - \mu Z^{-1} 1 - \Sigma^{-1} \Delta z
\end{aligned}
$$

$Z = \text{diag}(z)$; $S = \text{diag}(s)$

$$
= s - \mu Z^{-1} 1 - \Sigma^{-1} \Delta z
$$

and get.

$$\begin{bmatrix} H_x & J_E' & J_I \\ J_E & 0 & 0 \\ J_I & 0 & -\Sigma^{-1} \end{bmatrix} \begin{matrix} \Delta x \\ -\Delta y \\ -\Delta z \end{matrix} = \begin{matrix} \nabla f - J_E' y - J_I' z \\ c_E \\ c_I - \mu Z^{-1} 1 \end{matrix}$$

we could then eliminate Δz, too to get

$$\begin{bmatrix} H_x + J_I' \Sigma J_I & J_E' \\ J_I & 0 \end{bmatrix} \begin{matrix} \Delta x \\ -\Delta y \end{matrix} = \begin{matrix} \vdots \\ \vdots \end{matrix}$$

- may get fill in in top left corner.

- Each of these is ill conditioned, because some of $\Sigma \to \infty$ as $\mu \to 0$

- nonetheless, stable direct factorization methods work because of special form.

- Barrier parameter.

  - $\mu_{k+1} = \sigma_k \mu_k$

    options – fixed constant

    – two values:
    - small (last step made progress)
    - large (didn't)

  - $\mu_{k+1} = \sigma_k \dfrac{s_k' z_k}{n_I}$

Complementarity, from LP

Non-convexity, Singularity:

- step may not be productive, because it heads only to KKT points

- might get local <u>max</u> of f.

- if Hessian is pd on null space of constraints, then, for <u>equality</u> constraind probs, we get a descent step.

- for this case, we can show that if there are

$n + m$ +ve evals
(m = # ineqs)

$l + m$ −ve evals
(l = # eqs)

0 0 evals

in P.D. matrix we are OK.

- Make matrix mods while factoring to ensure this applies.

Accepting steps:

merit function:

$$f(x) - \mu \sum_i \log s_i + \nu \| C_E \| + \nu \| C_I - s \|$$

l_1, l_2 norm, not squared.

- find step; obtain α_s max, α_z max
- now search $0 \cdots \alpha_{s\,max}$

  $0 \cdots \alpha_{z\,max}$

  for best value of $\phi(x, s)$

~~Opt~~ Observations

- can do BFGS

  (replace Hessian w/ BFGS est)

- can ensure strict feasibility.

- can do trust region methods.

- can demonstrate quite strong convergence properties for trust region methods

- can obtain superlinear convergence ~~of~~ usually only close to a soln.