Interior point methods for QP

$$(QP)\begin{cases}\min \frac{1}{2}x^TQx + d^Tx \\ \text{s.t. } Ax \geq b\end{cases}$$

(equality constraints: ex for HW4)

$Q \in \mathbb{R}^{n\times n}$ symm pos ~~semidef~~

$A \in \mathbb{R}^{m\times n}$

KKT conditions:

$$\begin{cases} Qx - A^Tz + d = 0 \\ Ax - b \geq 0 \\ (Ax-b)_i z_i = 0 \\ z \geq 0 \end{cases}$$

Introduce slacks $y \geq 0$

$$\Leftrightarrow \quad (KKT)\begin{cases} Qx - A^Tz + d = 0 \\ Ax - b - y = 0 \\ y_i z_i = 0 \\ y \geq 0,\ z \geq 0 \end{cases}$$

$Y = \text{diag}(y_1, y_2 \dots y_m)$

$Z = \text{diag}(z_1, z_2 \dots z_m)$

$e = \begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix} \in \mathbb{R}^m$

Q pos semidef $\Rightarrow$ (KKT) is necessary and sufficient

To find solution to (QP) we need to solve (KKT), however we consider a slightly perturbed version:

$$(*) \quad F_\mu(x,y,z) = \begin{bmatrix} Qx - A^Tz + d \\ Ax - y - b \\ YZe - \mu e \end{bmatrix} = 0$$

, in order to guide iterations with the central path: $\{(x_\mu, y_\mu, z_\mu) \text{ sol to } (*)\}$

By applying Newton's method, we need to solve for step $(\Delta x, \Delta y, \Delta z)$

$$\underbrace{\begin{bmatrix} Q & 0 & -A^T \\ A & -I & 0 \\ 0 & Z & Y \end{bmatrix}}_{F'_\mu(x,y,z)} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \underbrace{\begin{bmatrix} -Q(x - A^Tz + d) \\ -(Ax - y - b) \\ -YZe + \mu e \end{bmatrix}}_{-F_\mu(x,y,z)}$$

At the solution to (QP) $YZe = 0$, so the smaller μ is the closer we are to the solution.

However we may be far from having $YZe = \mu e$, so we estimate progress to solution with duality gap $\frac{y^Tz}{m}$ = average duality gap

Thus the IPM for QP looks as follows.

Given (x_0, y_0, z_0) s.t. $y_0 > 0$ and $z_0 > 0$.
((x_0, y_0, z_0) need not be feasible)

for $k = 0, 1, \ldots$

$$\mu = \frac{y^T z}{m}$$

solve
$$\begin{bmatrix} Q & 0 & -A^T \\ A & -I & 0 \\ 0 & Z & Y \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} -Q(x_k - A^T z_k + d) \\ -(Ax_k - y_k - b) \\ -Y_k Z_k e + \mu \sigma e \end{bmatrix} \quad (\star)$$

$\sigma \in (0,1)$ makes $\mu_k \to 0$ & follow central path.

$$(x_{k+1}, y_{k+1}, z_{k+1}) = (x_k, y_k, z_k) + \alpha (\Delta x, \Delta y, \Delta z)$$

where α is chosen to maintain $y > 0$, $z > 0$
(possibly a fraction of the largest such α see § 16.6)

The systems $(\star)$ can be solved by essentially doing Gaussian elimination.

Eliminating Δy:

$$A\Delta x - \Delta y = -(Ax - y - b)$$

$$\Rightarrow \begin{bmatrix} Q & -A^T \\ A & Z^{-1}Y \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \end{bmatrix} = \begin{bmatrix} -(Qx - A^T z + d) \\ -(Ax - y - b) - y + \sigma \mu Z^{-1} e \end{bmatrix}$$

Similarly eliminating Δz:

$$(Q + A^T Y^{-1} Z A)\Delta x = -(Qx - A^T z + d) + A^T Y^{-1} Z \begin{bmatrix} -(Ax - y - b) \\ -y \\ +\sigma \mu Z^{-1} e \end{bmatrix}$$

which is a symmetric system that can be solved e.g. w/ Cholesky or CG.

§16.7 Bound constrained optimization problems

$$\min f(x) \quad \text{s.t. } a \leq x \leq b$$

(of course $a_i < b_i$ to have a feasible set)
$f: \mathbb{R}^n \to \mathbb{R}$ twice cont diff'ble
$a_i = -\infty$, $b_i = +\infty$ allowed.

This a generalization of what appears in (N & W QP).

Optimality conditions

$C = \{x \in \mathbb{R}^n \mid a \leq x \leq b\}$ = closed & convex

If x^* is a local sol to (1) then:

1) $\nabla f(x^*)^T (x - x^*) \geq 0$

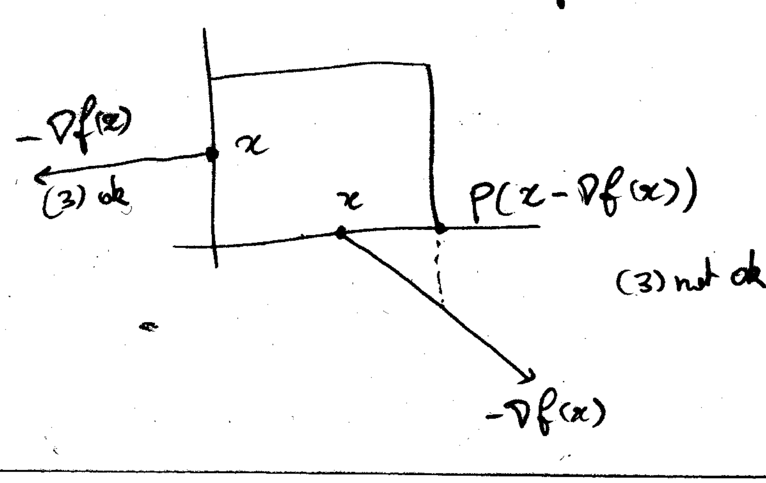
2) $\exists\ \lambda, \mu \geq 0$ s.t.

$$\nabla f(x^*) - \lambda + \mu = 0$$
$$\lambda^T (x^* - a) = 0$$
$$\mu^T (b - x^*) = 0$$

3) $P(x_* - \alpha \nabla f(x_*)) = x_* \quad \forall \alpha > 0$

Here P = projection onto C

$$(Px)_i = \begin{cases} a_i & \text{if } x_i < a_i \\ x_i & \text{if } a_i \leq x_i \leq b_i \\ b_i & \text{if } x_i > b_i \end{cases}$$



proof (3):

By (2): $(\nabla f(x))_i \geq 0$ if $(x_*)_i = a_i$

$(\nabla f(x))_i \leq 0$ if $(x_*)_i = b_i$

$(\nabla f(x))_i = 0$ if $(x_*)_i \in (a_i, b_i)$

So for $x^* \in C$, $\nabla f(x_*) = 0$

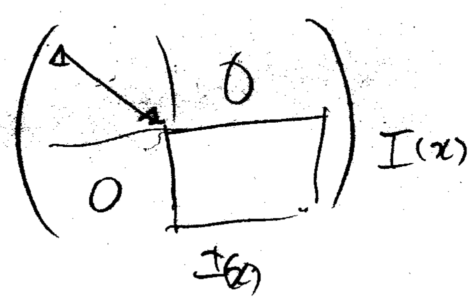
Def: $\mathcal{A}(x) = \{ i \mid a_i = x_i \text{ or } b_i = x_i \}$

(active indices)

$I(x) = \mathcal{A}^c(x)$ = inactive indices

Def: Reduced Hessian

$$\nabla_R^2 f(x) = \begin{cases} \delta_{ij} & \text{if } i \in \mathcal{A}(x) \text{ or } j \in \mathcal{A}(x) \\ (\nabla^2 f(x))_{ij} & \text{otherwise} \end{cases}$$

$=$ 

Second order nec. optimality conditions

If x^* is a local sol of (1) then $\nabla_R^2 f(x)$ is pos semidef

Def A point $x \in C$ is a nondegenerate stationary point if:

- $P(x_* - \alpha \nabla f(x_*)) = x^*$ $\forall \alpha \geq 0$

(⇔ strict complementarity)

- $(\nabla f(x_*))_i \neq 0$ then $i \in \mathcal{A}(x_*)$

Lemma Let x^* be a nondegenerate stationary point then $\exists\, \sigma > 0$ s.t.

$$\forall x \in C \quad \nabla f(x^*)^T (x - x^*) = \nabla f(x^*)^T I_{A^*} (x - x^*)$$
$$\geq \sigma \| I_{A^*} (x - x^*) \|_2$$

where $I_{A^*}(y) = \begin{cases} y_i & , i \in \mathcal{A}(x^*) \\ 0 & \text{otherwise} \end{cases}$

proof: x^* is a non-degenerate stat pt:

$\Rightarrow \exists\, \sigma > 0$ s.t.

$$|(\nabla f(x^*))_i| \geq \sigma, \quad \forall i \in \mathcal{A}(x^*)$$

if $x \in C$ $(\nabla f(x^*))_i (x - x^*)_i = \begin{cases} 0 & \text{if } i \in \overset{\text{since } \nabla f = 0}{I(x^*)} \\ (\nabla f(x))_i (x - x^*) & i \in \mathcal{A}(x^*) \end{cases}$

$$\geq \sigma |x_i - (x^*)_i|, \quad i \in \mathcal{A}(x^*)$$

to conclude sum over i and use $\|y\|_1 \geq \|y\|_2$.

Lemma says at least how much function value goes down if we go along active gradient components.

Theorem: If x^* is a non-degenerate stationary point If $\nabla_R^2 f(x^*)$ is pos def then x^* is a local min

proof: Let $x \in C$. $\quad \phi(t) = f(x^* + t(x - x^*))$

$$\phi'(0) = \nabla f(x^*)^T (x - x^*)$$
$$= \nabla f(x^*)^T I_{A^*} (x - x^*)$$
$$\cancel{+ \nabla f(x^*)^T I_{I^*} (x - x^*)} = 0$$

$\Rightarrow \phi'(0) > 0$ if $I_{A*}(x - x_*) \neq 0$.

If $I_{A*}(x - x_*) = 0$

$$\phi''(0) = (x - x_*)^T \nabla^2 f(x_*) (x - x_*)$$
$$= I_{I*}(x - x_*)^T \nabla^2 f(x_*) I_{I*}(x - x_*) > 0.$$

(which is what we assumed to be pos def).

Gradient projection algorithm (easy to implement)

Given x_c

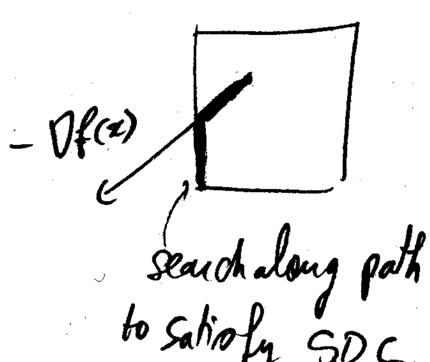
$x_+ = x_c(\alpha) = P(x_c - \alpha \nabla f(x_c))$, where $\alpha = \left(\frac{1}{2}\right)^m$ and m is smallest integer s.t. Sufficient decrease condition is satisfied

$$f(x_c(\alpha)) - f(x_c) \leq \frac{\sigma}{\alpha} \|x_c - x_c(\alpha)\|_2^2$$

if $C = \mathbb{R}^n$ then exactly SDC for unconstrained optil

take $\sigma \sim 10^{-4}$.

This "line search" is easy to implement and amounts to doing line search on piecewise linear paths:

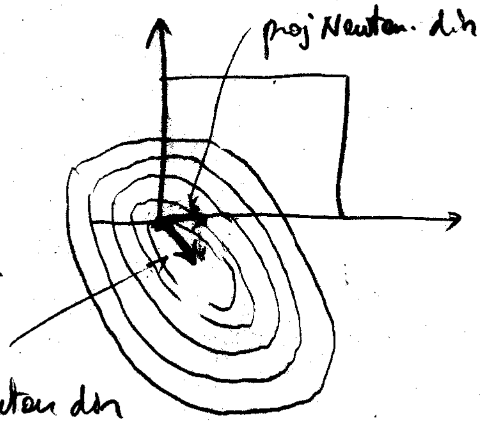


Can prove that if x_* is non degenerate, $A(x_k) = A(x_*)$ for k sufficiently large

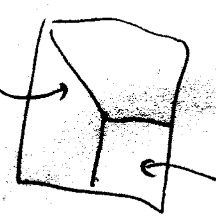
Gradient projection ~ Gradient method with constraints aware line search.

Can we use Newton's direction instead?

- Projected Newton's direction is not a descent dir for this constrained problem.



<u>Solution</u>: used reduced Hessian instead



~ hybrid method.