Trust region SQP methods

$$\begin{cases} \min & f(x) \\ \text{s.t.} & c(x) = 0 \end{cases}$$

Let x_k, λ_k be current iterate and Lagrange multipliers.

The TR subproblem for SQP is:

$$(1) \begin{cases} \min & \nabla f_k^T s + \frac{1}{2} s^T H_k s \\ \text{s.t.} & c'_k s + c_k = 0 \\ & \|s\|_2 \leq \Delta_k \end{cases}$$

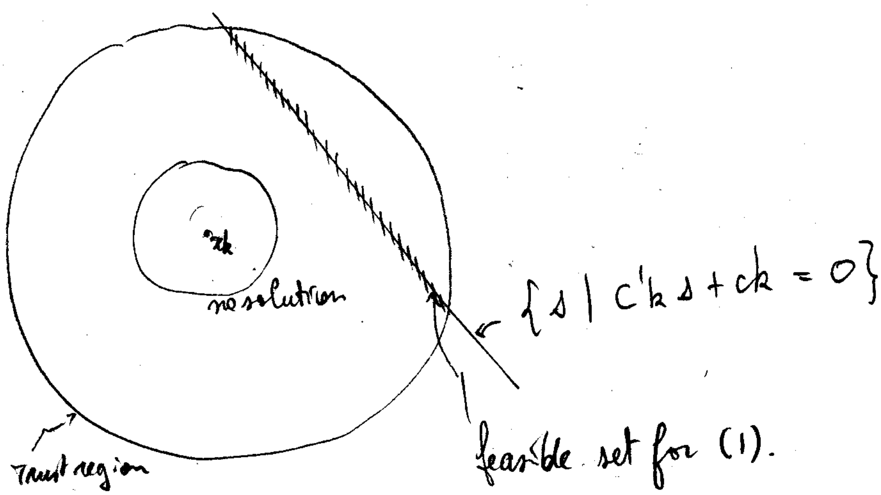
$\leftarrow$ only difference w.r.t unconstrained TR method.

Recall H_k is a replacement of $\nabla_{xx} L(x_k, \lambda_k)$

Δ_k = TR radius

$c'_k s + c_k = 0 \iff$ linearized constraints.

However there is no guarantee that (1) has a solution.



A frequently used way of getting rid of this artifact introduced by QP is to decouple the problem:

$$s = s^t + s^n$$

where s_n = quasi normal step (or vertical) which moves towards feasibility

s_t = tangential step which moves towards optimality (a.k.a horizontal)

s^n is an approx sol of:

$$(2)\ \begin{cases} \min \frac{1}{2} \|C'_k s^n + C_k\|_2^2 = q(s^n) \\ \text{s.t. } \|s^n\|_2 \leq \theta \Delta_k \end{cases}, \text{ where } \theta \in (0,1) \text{ is a fixed param.}$$

Requirement on s^n for convergence:

- $\|s^n\| \leq \theta \Delta_k$
- $q(s^n)$ must decrease at least a fraction of Cauchy decrease:

$$q(0) - q(s^n) \geq \underset{\substack{\cap \\ (0,1)}}{\sigma} \left(q(0) - q(s^{CP}) \right)$$

$$\Leftrightarrow \quad \|C_k\|_2^2 - \|C'_k s^n + C_k\|_2^2 \geq \sigma \left(\|C_k\|_2^2 - \|C'_k s^{CP} + C_k\|_2 \right)$$

recall $s^{CP} = -t^{CP} C'^T_k C_k$ $\quad \left\{ q(s) = \frac{1}{2}\|C_k\|^2 + \underbrace{(C'^T_k C_k)^T}_{\text{linear part}} s + \frac{1}{2} s^T \underbrace{C'^T_k C'_k}_{\text{Hessian}} s \right.$

Cauchy point for (2)

$$t^{CP} = \begin{cases} \dfrac{\|C'^T_k C_k\|_2^2}{(C'^T_k C_k)(C'^T_k C'_k)(C'^T_k C_k)} & \text{if } \dfrac{\|C'^T_k C_k\|_2^3}{(\text{same denominator})} \leq \theta \Delta_k \text{ and (denominator)} > 0 \\ \dfrac{\Delta_k}{\|C'^T_k C_k\|_2} & \text{otherwise} \end{cases}$$

- since $C'_k = {}_m\overset{n}{\boxed{\qquad\qquad}}$ $(m < n)$, we will have many solutions

s_n to (2). We could take minimal norm sol (as suggested in book), but it is enough to enforce:

$$\|s^n\| \leq \underset{\substack{\uparrow \\ \text{some pos. const.}}}{K} \|C_k\|$$

Then we look for $s = s^n + s^t$ as an approx sol to:

not zero but relaxed to what we can achieve with s_n.

$$(3) \quad \begin{cases} \min & \nabla f_k^T s + \frac{1}{2} s^T H_k s \\ \text{s.t.} & C'_k s + c_k = C'_k s_n + c_k \\ & \|s\| \leq \Delta_k \end{cases}$$

Since $s = s^n + s^t$, the constraints become:

$$C'_k s^t = 0.$$

$\Rightarrow$ Tangential step $\in \mathcal{N}(C'_k)$.

Thus (3) is equivalent to:

$$(4) \quad \begin{cases} \min & (\nabla f_k + H_k s^n)^T s^t + \frac{1}{2} s^{t\,T} H_k s^t \\ \text{s.t.} & C'_k s^t = 0 \\ & \|s^t + s^n\| \leq \Delta_k \end{cases}$$

Let Z_k be a representation of ~~nullspace of~~ lin. constraints i.e.:

$$\mathcal{R}(Z_k) = \mathcal{N}(C'_k), \quad Z_k \in \mathbb{R}^{n \times (n-m)}$$

Then:

$$\underset{\mathbb{R}^n}{s^t} = Z_k \underset{\mathbb{R}^{n-m}}{\hat{s}}$$

(typically $\hat{s}$ is a much smaller vector)

Then (4) is equivalent to:

$$\begin{cases} \min & \left[Z_k^T (\nabla f_k + H_k s^n)\right]^T \hat{s} + \frac{1}{2} \hat{s}^T Z_k^T H_k Z_k \hat{s} = q \\ \text{s.t.} & \|Z_k \hat{s} + s^n\| \leq \Delta_k \end{cases}$$

Trust region constraints have a shift. Can do change of variables.

In order to get convergence $q(\hat{s})$ should satisfy a fraction of Cauchy decrease