SQP + QP + Active set

https://www.youtube.com/watch?v=m5Gos-fh9hc

Original Problem

$$\min_{x \in \mathbb{R}^n} f(x) \quad \Rightarrow \quad L(x,\lambda,\mu) := f(x) + \lambda^T h(x) + \mu^T g(x)$$

$$\text{s.t.} \quad g(x) \le 0 \in \mathbb{R}^m, \quad h(x) = 0 \in \mathbb{R}^p$$

$\Downarrow$

KKT ①:

$$\nabla_x L(x^*,\lambda,\mu) = \nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0$$
$$\nabla_\lambda L(x^*,\lambda,\mu) = h(x^*) = 0$$
$$\nabla_\mu L(x^*,\lambda,\mu) = g(x^*) \le 0$$
$$\mu^T g(x^*) = 0$$

If: $g(x)$ not exist $\Leftrightarrow$ equality constraint only $\Leftrightarrow$ no complementarity constraint

$\Rightarrow$ Newton or Quasi-Newton method (SQP)

Newton-SQP step to find root of KKT condition: $\Delta x = -\frac{f(x)}{\nabla f(x)}$, here $f(x) = \nabla L$

$$\underbrace{\nabla^2_{(x,\lambda)} L(x_k,\lambda_k)}_{\shortparallel} \begin{pmatrix} d \\ \delta \end{pmatrix} = -\nabla_{(x,\lambda)} L(x_k,\lambda_k)$$

$$\begin{bmatrix} \nabla^2_{xx} L(x_k,\lambda_k) & \nabla h_1(x_k) \cdots \nabla h_p(x_k) \\ \nabla h_1^T(x_k) & \\ \vdots & 0 \\ \nabla h_p^T(x_k) & \end{bmatrix}$$

or this is also derived from Taylor expansion $L(x,\lambda)$ and then find its minimum

$$L(x_k+d, \lambda_k+\delta) = L(x_k,\lambda_k) + \nabla^T L \cdot \begin{pmatrix} d \\ \delta \end{pmatrix} + \frac{1}{2}\nabla^2 L \begin{pmatrix} d \\ \delta \end{pmatrix}^2 := L_k(d,\delta), \text{ fix } x_k, \lambda_k$$

$$\therefore \nabla L(\quad) = 0 \Rightarrow \nabla^2 L \cdot \begin{pmatrix} d \\ \delta \end{pmatrix} = -\nabla L$$

$$\therefore L_k(d,\delta) = \frac{1}{2} d^T \cdot \nabla^2_{xx} L \cdot d + \nabla_x L^T d + L(x_k,\lambda_k) + \underset{\text{new multiplier}}{\delta^T} \nabla^2_{x\lambda} L d + \delta^T h(x_k)$$

is the Lagrangian of:

$$\min_{d \in \mathbb{R}^n} \frac{1}{2} d^T \nabla^2_{xx} L \cdot d + \nabla_x L^T d + L(x_k,\lambda_k)$$
$$\text{s.t.} \quad \nabla^2_{x\lambda} L d + h(x_k) = 0$$

$\Rightarrow$ QP, done!

Now with Inequality $\quad L(x_k+d, \lambda_k+\delta, \mu_k+y) \approx L + \nabla_{x\lambda\mu} L^T \begin{pmatrix} d \\ \delta \\ y \end{pmatrix} + \frac{1}{2} \nabla^2_{x\lambda\mu} \begin{pmatrix} d \\ \delta \\ y \end{pmatrix}^2 =: L_k(d, \delta, y)$

Use Newton is not possible with inequality here.

$$L_k(d, \delta, y) = \frac{1}{2} d^T \nabla^2_{xx} L \cdot d + \nabla_x L^T d + L + \delta^T \underbrace{\nabla^2_{x\lambda} L}_{= (\nabla h_i^T(x) \cdot d)_{i=1}^p} \cdot d + y^T \underbrace{\nabla^2_{x\mu} L}_{= (\nabla g_i^T(x) \cdot d)_{i=1}^m} \cdot y + \delta^T h(x_k) + y^T g(x_k)$$

which is the Lagrangian of:

$$\min_{d \in \mathbb{R}^n} \frac{1}{2} d^T \nabla^2_{xx} L \cdot d + \nabla_x L^T d + L$$

$$\text{s.t.} \left.\begin{cases} h_i(x_k) + \nabla h_i(x_k)^T \cdot d = 0, & i = 1, \cdots p \\ g_i(x_k) + \nabla g_i(x_k)^T d \le 0, & i = 1, \cdots m \end{cases}\right\} \text{Linearized constraint}$$

$$\Longleftrightarrow \quad \min_{d \in \mathbb{R}^n} \frac{1}{2} d^T A d + b^T \cdot d + c, \quad \text{s.t.} \begin{cases} Hd + h = 0 \\ Gd + g \le 0 \end{cases}$$

A simpler QP with linear constraint!

Now construct yet another Lagrangian for QP:

$$L(d, \lambda, \mu) = \frac{1}{2} d^T A d + b^T \cdot d + c + \lambda^T (Hd + h) + \mu^T (Gd + g)$$

and KKT 2

$$\begin{aligned} Ad + b + H^T \lambda + G^T \mu &= 0 \\ Hd + h &= 0 \\ Gd + g &\le 0 \\ \mu^T (Gd + g) &= 0 \end{aligned}$$

Active set: at admissible d, we have active constraints $\mathcal{A} = \{j \in \{1, \cdots m\}, (Gd)_j + g_j) = 0\}$

and inactive constraints $I = \{j \in \{1, \cdots m\}, (Gd)_j + g_j) < 0\}$

property: $\mu_j = 0, \ j \in I$

New KKT 3:

(homogeneous problem)

$$\begin{aligned} A\hat{d} + H^T \lambda + \hat{G}^T \cdot \hat{\mu} &= -(Ad + b) \\ H \hat{d} &= 0 \\ \hat{G} \hat{d} &= 0 \end{aligned}$$

$\hat{G}$ contains rows of active constrainsts if the solution of 3 is $(\hat{d}, \lambda, \hat{\mu})$. set. $\mu_{i\in A} = \hat{\mu}$ and $\mu_{i\in I} = 0$.

then $\hat{G}\hat{\mu} = G\mu$, if $\mu \geq 0$, then $(d+\hat{d}, \lambda, \mu)$ solves the KKT 2

$d+\hat{d}$ is admissible! then we can update d, active set, and iterate forward.

Algorithm: Active set strategy

1. Find an admissible point $d \in \mathbb{R}^n$.
2. For d, determine the set $\mathcal{A}$ of active inequality constraints defined in (1).
3. Compute a solution $(\hat{d}, \lambda, \hat{\mu})$ of the linear system (2).

(a) If $\|\hat{d}\| \leq \epsilon$:
- If $\hat{\mu} \geq 0$: stop, (d, λ, μ) with $(\mu_i)_{i\in\mathcal{A}} := \hat{\mu}, (\mu_i)_{i\in\mathcal{I}} := 0$ solves the KKT system.
- If there is $i \in \mathcal{A}$ with $\hat{\mu}_i < 0$: set $\mathcal{A} = \mathcal{A}\backslash\{i\}$ and go back to step 3.
  (Sensitivity Theorem: $\hat{\mu}_i < 0$: $(\hat{G}\hat{d})_i < 0 \Rightarrow f \downarrow$, inactive constraint will reduce cost.)

(b) If $\|\hat{d}\| > \epsilon$:
- Choose step-size ρ and set $d := d + \rho\hat{d}$.
  ($H\hat{d} = 0, \hat{G}\hat{d} = 0 \rightsquigarrow$ this will not violate the equality and the active inequality constraints.)
  Inactive constraints ($i \in \mathcal{I}$) must not be violated, too:

$$(G(d+\rho\hat{d})+g)_i = \underbrace{(Gd)_i + g_i}_{\leq 0} + \rho(G\hat{d})_i \leq 0, i \in \mathcal{I} \quad \Rightarrow \quad \rho = \min_{i\in\mathcal{I}}\left\{\frac{(Gd)_i + g_i}{(G\hat{d})_i}\right\}.$$

- Go back to step 2.