Sequential Quadratic Programming

Note Title 5/28/2022

SQP for constrained optimization problems is like Newton/Quasi-Newton methods for unconstrained problems.

$\min f(x)$ st. $c_i(x)=0,\ i\in\mathcal{E}$
$c_i(x)\geq 0,\ i\in\mathcal{I}$

"quadratic model"

$\approx \underbrace{\min q_k(x) \text{ st } l_i^k(x)=0,\ i\in\mathcal{E},\ l_i^k(x)\geq 0,\ i\in\mathcal{I}}_{\text{some QP approx. near } x_k}$ $\rightarrow x_{k+1}$, and iterate

$\min f(x)$

$\approx \min q_k(x)$ $\rightarrow x_{k+1}$, and iterate

some quadratic approx. of f near x_k

The iterates are not required to be feasible. (In general, finding a feasible point when the constraints are nonlinear may be as hard as solving the optimization problem itself.)

How to pick the QP subproblem?

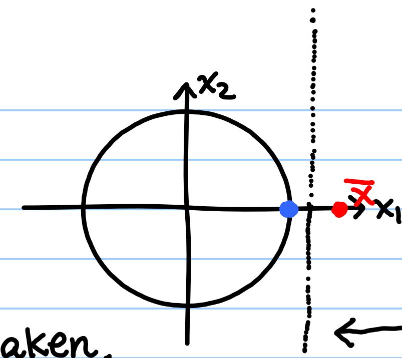
The most obvious choice is: $q_k(x)=f_k+\nabla f_k(x-x_k)+\frac{1}{2}(x-x_k)^T\nabla^2 f_k(x-x_k)$ (local quad. approx.)

$l_i^k(x)=c_i^k+\nabla c_i^k(x-x_k)$ (local linear approx.)

The following example shows that this choice does not work even locally when the c_i's are nonlinear.

nonconvex quadratic   nonlinear

Consider $\min_x \; -x_1 - \frac{1}{2}x_2^2 \quad \text{s.t. } x_1^2 + x_2^2 - 1 = 0$



Global solution at $[1,0]^T$.

It would not matter where the local approximation ($\bar{x}$) is taken, the resulted QP is unbounded.

E.g. If $\bar{x} = \begin{bmatrix} 1+\varepsilon \\ 0 \end{bmatrix}$, the QP is $\min_x \; -x_1 - \frac{1}{2}x_2^2$ s.t. $c(\bar{x}) + 2\bar{x}^T(x-\bar{x}) = 0$

$$\begin{aligned} & \;\; \| \\ & (1+\varepsilon)^2 - 1 + 2(1+\varepsilon)(x_1 - (1+\varepsilon)) \end{aligned}$$

$$\Leftrightarrow \; x_1 - (1+\varepsilon) = -\varepsilon(2+\varepsilon)/2(1+\varepsilon)$$

Since x_2 can be any value in the linearized constraint, the QP is unbounded.

Here $\tilde{x} \approx \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, meaning that there is no hope to obtain even a local convergence result with the simple quadratic model.

Interestingly, the following quadratic model works a lot better:

$$(*)\begin{cases} \min_x f(x) & \\ \text{s.t.} \quad c_i(x) = 0 & i \in \mathcal{E} \\ \qquad\ c_i(x) \geq 0 & i \in \mathcal{I} \end{cases} \approx \left.\begin{array}{lll} \min_p & f_k + \nabla f_k^T p + \frac{1}{2} p^T \nabla_{xx}^2 \mathcal{L}_k p^T & \\ \text{s.t.} & c_i(x_k) + \nabla c_i(x_k)^T p = 0 & i \in \mathcal{E} \\ & c_i(x_k) + \nabla c_i(x_k)^T p \geq 0 & i \in \mathcal{I} \end{array}\right\} \star$$

$$\left[\mathcal{L}(x,\lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x), \quad \nabla_{xx}^2 \mathcal{L}_k = \nabla_{xx}^2 \mathcal{L}(x_k, \lambda_k)\right]$$

Why this quadratic model?

It's easy to see for equality-constrained problems:

$$\boxed{\text{When } \mathcal{I} = \emptyset, \text{ and under suitable technical conditions, solving } \star \text{ is the same as one step of the (pure) Newton's method applied to the system of nonlinear equations given by the KKT conditions of } (*).}$$

So when $\mathcal{I} = \phi$ the SQP method based on $\star$ is exactly the Newton's method applied to the KKT system. This means that when (x_0, λ_0) is close enough to a KKT point (x^*, λ^*) not only would the SQP method produces a sequence $\{x_k\}$ s.t. $x_k \to x^*$, it converges very rapidly:

$$\|x_{k+1} - x^*\| \leq C \|x_k - x^*\|^2. \qquad \text{(Quadratic convergence)}$$

Let's prove the claim above. (After the proof, we discuss the case of $\mathcal{I} \neq \emptyset$.)

The KKT conditions of (*) (when $\mathcal{I} = \emptyset$) is

$$F(x,\lambda) = \begin{bmatrix} \nabla f(x) - A(x)^T \lambda \\ c(x) \end{bmatrix} = 0, \qquad \underset{n\times m}{A(x)^T} = [\nabla c_1(x), \cdots, \nabla c_{\overset{\downarrow}{m}}(x)]. \quad m = |\mathcal{E}|$$

Assume (x^*, λ^*) is a KKT point (i.e. $F(x^*, \lambda^*) = 0$).

The Newton step from the iterate (x_k, λ_k) is $\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix} + \begin{bmatrix} p_k \\ l_k \end{bmatrix}$, where p_k, l_k solve the linear system:

$$\underbrace{\underset{DF(x_k,\lambda_k)=}{\begin{bmatrix} \nabla^2_{xx} \mathcal{L}_k & -A_k^T \\ A_k & 0 \end{bmatrix}} \begin{bmatrix} p_k \\ l_k \end{bmatrix} = \underset{=-F(x_k,\lambda_k)}{\begin{bmatrix} -\nabla f_k + A_k^T \lambda_k \\ -c_k \end{bmatrix}}}$$

$\Updownarrow$ Ex: Check it.

$$\boxed{\begin{bmatrix} \nabla^2_{xx} \mathcal{L}_k & -A_k^T \\ A_k & 0 \end{bmatrix} \begin{bmatrix} p_k \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} -\nabla f_k \\ -c_k \end{bmatrix} \qquad \text{(Newton-KKT)}}$$

directly solving λ_{k+1}, instead of solving the update l_k.

The KKT conditions of the QP in ⊛ (when $\mathcal{I} = \emptyset$) is exactly the same as that of (Newton-KKT).

Ex: Check it. (Put differently, the QP in ⊛ is constructed so that this property holds)

Technicality: If we assume $DF(x^*, \lambda^*)$ is non-singular, then so is $DF(x_k, \lambda_k)$ when $(x_k, \lambda_k) \approx (x^*, \lambda^*)$.

$$DF(x_k, \lambda_k) = \begin{bmatrix} \nabla^2_{xx} \mathcal{L}_k & -A_k^T \\ A_k & 0 \end{bmatrix}$$

The non-singularity of $DF(x^*, \lambda^*)$, in turn, is implied by the following (familar) conditions:

(i) $A(x^*)$ has full row rank (the LICQ condition)

+

(ii) $\nabla^2_{xx} \mathcal{L}(x^*, \lambda^*)$ is pos. def. on the tangent space of the constraints, i.e.

$$d^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) d > 0 \quad \text{for all } d \neq 0 \text{ s.t. } A(x^*) d = 0$$

($\Leftrightarrow$ $Z^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z$ is positive definite when $\underset{m \times n}{A(x^*)} \underset{n \times (n-m)}{Z} = 0$, $\operatorname{rank}(Z) = n-m$.)

(2nd order sufficient condition for optimality.)

When (i) + (ii) holds and $(x_k, \lambda_k) \approx (x^*, \lambda^*)$, we also have $A(x_k)$ is full rank and
$d^T \nabla^2_{xx} \mathcal{L}(x_k, \lambda_k) d > 0 \quad \forall\, d \neq 0,\ A(x_k) d = 0.$

In this case, by the first result in the QP chapter, we also conclude that the unique solution of (Newton-KKT) is the unique global solution of the QP in ⊛.

So the claim is proved (under conditions (i) and (ii).) Q.E.D.

So the new iterate (x_{k+1}, λ_{k+1}) can therefore be defined either as the solution of the QP

$$\begin{aligned} \min_p \ & f_k + \nabla f_k^T p + \tfrac{1}{2} p^T \nabla^2_{xx} \mathcal{L}_k p^T \qquad (18.7) \quad (= ⊛ \text{ with } \mathcal{I} = \phi.) \\ \text{s.t.} \ & c_i(x_k) + \nabla c_i(x_k)^T p = 0 \quad i \in \mathcal{E}, \end{aligned}$$

or

as the iterate generated by Newton's method applied to the KKT conditions of the original equality-constrained problem.

Why bother to interpret the same method in two different ways?

- The Newton point of view facilitates the analysis.
- The SQP framework enables us to derive practical algorithms and to extend the technique to the inequality-constrained case.

Why would the SQP method based on ⊛ also work locally when $\mathcal{I} \neq \emptyset$?

$$(*) \left\{ \begin{array}{lll} \min_x f(x) & & \\ \text{s.t.} & c_i(x) = 0 & i \in \mathcal{E} \\ & c_i(x) \geq 0 & i \in \mathcal{I} \end{array} \right.$$

" (18.10)

x^* a local solution
active set $\mathcal{A}(x^*)$

Choose (x_0, λ_0).
for $k = 0, 1, 2, \cdots$

$$\left. \begin{array}{llll} \text{Solve} & \min_p f_k + \nabla f_k^T p + \frac{1}{2} p^T \nabla_{xx}^2 \mathcal{L}_k p^T & & \\ & \text{s.t.} \quad c_i(x_k) + \nabla c_i(x_k)^T p = 0 & i \in \mathcal{E} \\ & \phantom{\text{s.t.} \quad} c_i(x_k) + \nabla c_i(x_k)^T p \geq 0 & i \in \mathcal{I} \end{array} \right\} ⊛$$

" (18.11)

$\rightarrow$ new iterate $(x_{k+1} = x_k + p_k, \lambda_{k+1})$

active set $\mathcal{A}_k$

end

[local Newton-SQP method]

A key observation:

> If $\mathcal{A}(x^*) = \mathcal{A}_k$ for all large enough k, i.e. the SQP method is able to correctly identity the optimal active set, then the SQP method will act like a Newton method for equality-constrained optimization. Therefore, in virtue of our analysis in the case of $\mathcal{I} = \emptyset$, the SQP method will converge — and converge rapidly.

The following result gives conditions under which this desirable behavior takes place.

Theorem 18.1 (Robinson [267]).

Suppose that x^* is a local solution of (18.10) at which the KKT conditions are satisfied for some λ^*. Suppose, too, that the linear independence constraint qualification (LICQ) (Definition 12.4), the strict complementarity condition (Definition 12.5), and the second-order sufficient conditions (Theorem 12.6) hold at (x^*, λ^*). Then if (x_k, λ_k) is sufficiently close to (x^*, λ^*), there is a local solution of the subproblem (18.11) whose active set $\mathcal{A}_k$ is the same as the active set $\mathcal{A}(x^*)$ of the nonlinear program (18.10) at x^*.

Recall : Strict complimentarity is said to hold at a solution pair (x^*, λ^*) if

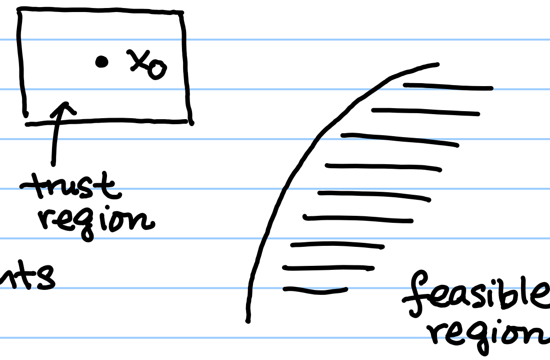
$$\nexists\, i \in \mathcal{I} \text{ s.t. } \lambda_i^* = c_i(x^*) = 0.$$

Even if a KKT point (x^*, λ^*) possesses all three properties above, when (x_0, λ_0) is far from (x^*, λ^*), a litany of problems may occur to the local Newton-SQP method:

- the QP subproblem is not convex anymore. (This is already a problem for unconstrained problems. Recall that we had several techniques to deal with it : modification of the Hessian, adding a trust region constraint, using quasi-Newton methods.)

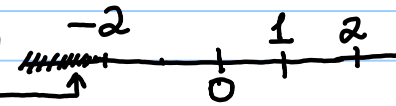
- Each of these three methods, when adapted to constrained problems, will face its own challenge.

  E.g. adding a trust region may render the QP subproblem infeasible.



- inconsistent linearizations:
  if $x_0 \notin$ feasible region, the linearized constraints may be inconsistent.

E.g. Consider $\{x \in \mathbb{R}^1 : x \leq 1,\ x^2 \geq 4\} =: S \neq \emptyset$



Let's linearize the two constraints at $x_0 = 1 \notin S$:

$$\left.\begin{array}{lll} 1-x \geq 0 & \Longleftrightarrow \quad 1-x = 0-1\overbrace{(x-1)}^{p} \geq 1 & \Leftrightarrow \quad p \leq -1 \\ x^2-4 \geq 0 & \xrightarrow[@x_0=1]{\text{linearize}} \quad -3+2\overbrace{(x-1)}^{p} \geq 0 & \Leftrightarrow \quad p \geq 3/2 \end{array}\right\} \text{inconsistent!}$$

- The Maratos effect