SI152: Numerical Optimization

Lecture 17: Sequential Quadratic Programming (SQP)

Hao Wang
Email: haw309@gmail.com

ShanghaiTech University

December 5, 2024

Robert B. Wilson (born May 16, 1937, Geneva, Nebraska; graduated from Harvard and work at Standford) is an American economist who was awarded the 2020 Nobel Prize for Economics for his contributions to the theory of auctions and for his invention of new auction formats, or rules of operation, for goods and services that could not be efficiently sold in.

SQP algorithm originates from his PhD dissertation:

"A Simplicial Algorithm for Concave Programming.

Boston: Harvard Business School, 1963".

Outline

1. Algorithmic Development
2. Algorithmic Enhancements (Equalities Only)
3. Inequality Constrained Problems

As for describing penalty methods, we begin with equality constrained problems:

$$\min_x \; f(x) \quad \text{s.t. } c(x) = 0$$

The Lagrangian for this problem is

$$L(x, \lambda) = f(x) + \lambda^T c(x),$$

and so the KKT conditions (necessary if $\nabla c(x)$ has full column rank) are

$$\begin{aligned} \nabla f(x) + \nabla c(x)\lambda &= 0 \\ c(x) &= 0. \end{aligned}$$

We interpret the central idea of SQO in two ways.

SQO subproblem: Interpretation 1

Let x_k be given. A local approximation of the constrained problem is

$$\begin{aligned} &\min_x \ f(x) \\ &\text{s.t. } c(x) = 0 \end{aligned} \quad \Longrightarrow \quad \begin{aligned} &\min_d \ f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T H_k d \\ &\text{s.t. } c(x_k) + \nabla c(x_k)^T d = 0 \end{aligned}$$

This is an equality constrained quadratic problem! (linear models and quadratic constrained models are also OK, but...)

- Constraints are affine, so KKT conditions for the subproblem are necessary.
- Assume $H_k \succeq 0$, so the KKT conditions are also suffcient.
- The Lagrangian for the subproblem is

$$L(d, \sigma) = f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T H_k d + \sigma^T (c(x_k) + \nabla c(x_k)^T d)$$

  and so the KKT conditions are

$$\begin{bmatrix} H_k & \nabla c(x_k) \\ \nabla c(x_k)^T & 0 \end{bmatrix} \begin{bmatrix} d \\ \sigma \end{bmatrix} = - \begin{bmatrix} \nabla f(x_k) \\ c(x_k) \end{bmatrix}.$$

- What should be the choice of H_k?

The KKT conditions for the subproblem on the previous slide, i.e.,

$$\begin{bmatrix} H_k & \nabla c(x_k) \\ \nabla c(x_k)^T & 0 \end{bmatrix} \begin{bmatrix} d \\ \sigma \end{bmatrix} = - \begin{bmatrix} \nabla f(x_k) \\ c(x_k) \end{bmatrix}.$$

should be familiar to anyone who has studied Newton's method for equations.

- Recall the optimality conditions for the nonlinear problem:

$$\begin{aligned} \nabla f(x) + \nabla c(x)\lambda &= 0 \\ c(x) &= 0 \end{aligned}$$

- At (x_k, λ_k), Newton's method for this system of equations computes

$$\begin{bmatrix} \nabla^2_{xx} L(x_k, \lambda_k) & \nabla c(x_k) \\ \nabla c(x_k)^T & 0 \end{bmatrix} \begin{bmatrix} d \\ \sigma \end{bmatrix} = - \begin{bmatrix} \nabla f(x_k) \\ c(x_k) \end{bmatrix}.$$

What is the difference? — with $H_k = \nabla^2_{xx} L(x_k, \lambda_k)$, they are equivalent.

Outline

The SQP subproblem is designed to achieve fast local convergence.

- Suppose we are in small neighborhood of a solution (x_*, λ_*) at which LICQ and the second order sufficient conditions hold. Then, for all k,
  - $\nabla c(x_k)^T$ has full row rank
  - $\nabla^2_{xx} L(x_k, \lambda_k)$ is positive definite in Null$(\nabla c(x_k)^T)$
  - (Recall that this means that the Newton primal-dual matrix is invertible.
- In this case, convergence follows from Newton's method — it is quadratic!
- However, what happens when we are not in the neighborhood of such a solution? We must have a globalization mechanism (e.g., line search, trust region, etc.).
- We must handle (local) nonconvexity.
- We must handle inconsistent linearizations of the constraints.

A given direction d_k may be

- descent direction for $\|c\|$ and descent direction for f;
- descent direction for $\|c\|$ and ascent direction for f;

Should we allow this? (Yes, we have to.) How do we judge progress?

- Need some mechanism for weighing progress in f and in c, e.g.,

  $$\phi(x;\nu) := f(x) + \nu\|c(x)\|_1$$

  Thus, we do NOT use $\phi(x)$ to compute the direction, but we DO use it as a tool for judging progress toward a solution of the nonlinear problem.
- How do we know that we have a descent direction for $\phi(x;\nu)$?

Theorem 1

Let (d_k, λ_{k+1}) be the computed SQP step. Then the directional derivative of the ℓ_1 merit function equals

$$D\phi(d_k; x_k, \nu) = \nabla f(x_k)^T d_k - \nu\|c(x_k)\|_1$$

and satisfies the bound

$$D\phi(d_k; x_k, \nu) \leq -d_k^T \nabla_{xx}^2 L(x_k, \lambda_k) d_k - (\nu - \|\lambda_{k+1}\|_\infty)\|c(x_k)\|_1.$$

SQP for equality constrained optimization

Choose initial (x_0, λ_0) and $\nu_{-1} > 0$
Choose $\eta \in (0,1)$ and some small $\epsilon > 0$.
for $k = 0, 1, 2, \ldots$ **do**
    Evaluate the SQP direction:

$$\begin{bmatrix} \nabla^2_{xx} L(x_k, \lambda_k) & \nabla c(x_k) \\ \nabla c(x_k)^T & 0 \end{bmatrix} \begin{bmatrix} d \\ \sigma \end{bmatrix} = - \begin{bmatrix} \nabla f(x_k) \\ c(x_k) \end{bmatrix}.$$

    Update the penalty parameter $\nu \leftarrow \max\{\nu, \|\lambda_{k+1}\|_\infty + \epsilon\}$.
    Choose $\alpha_k \in \{1, \frac{1}{2}, \frac{1}{4}, \ldots\}$ as the largest value such that

$$\phi(x_k + \alpha_k d_k; \nu) \leq \phi(x_k; \nu) + \eta \alpha_k D\phi(d_k; x_k, \nu).$$

    Update the iterate $x_{k+1} \leftarrow x_k + \alpha_k d_k$.
end for

If $\nabla^2_{xx} L(x_k, \lambda_k)$ is not positive definite in the null space of $\nabla c(x_k)^T$, then we will know since the inertia of the primal-dual matrix

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

may not be invertible.

If $\nabla c(x_k)^T$ has full row rank, this can be remedied with

$$\begin{bmatrix} \nabla^2_{xx} L(x_k, \lambda_k) + \mu I & \nabla c(x_k) \\ \nabla c(x_k)^T & 0 \end{bmatrix}$$

If $\nabla c(x_k)^T$ does not have full row rank, further remedied with

$$\begin{bmatrix} \nabla^2_{xx} L(x_k, \lambda_k) + \mu I & \nabla c(x_k) \\ \nabla c(x_k)^T & \delta I \end{bmatrix}$$

An alternative is to relax the constraints:

$$c(x_k) + \nabla c(x_k)^T d = r - s, \ (r, s) \geq 0$$

The SQO subproblem can be reformulated as

$$\min_d \ f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T H_k d + \nu \sum (r^i + s^i)$$
$$\text{s.t. } c(x_k) + \nabla c(x_k)^T d = r - s, \quad (r, s) \geq 0$$

How to choose ν? We have to solve an inequality constrained quadratic problem!

Outline

Consider the generally constrained nonlinear optimization problem:

$$\begin{aligned} \min_{x} \ & f(x) \\ \text{s.t. } & c^{\mathcal{E}}(x) = 0 \\ & c^{\mathcal{I}}(x) \le 0 \end{aligned}$$

A quadratic model of the objective and linearized constraints yield

$$\begin{aligned} \min_{d} \ & f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T \nabla^2_{xx} L(x_k, \lambda_k) d \\ \text{s.t. } & c^{\mathcal{E}}(x_k) + \nabla c^{\mathcal{E}}(x_k)^T d = 0 \\ & c^{\mathcal{I}}(x_k) + \nabla c^{\mathcal{I}}(x_k)^T d \le 0 \end{aligned}$$

We lose "Newton's method applied to the optimality conditions" interpretation. (... or do we?)

Theorem 2

Suppose that x_* is an optimal solution with corresponding KKT multipliers λ_* and that the following assumptions hold:

- *LICQ holds at x_**
- *Strictly complementarity holds at (x_*, λ_*), i.e.,*

$$\lambda_*^i + c^i(x_*) > 0, \quad i \in \mathcal{A}(x_*).$$

- *Second order sufficient condition holds at (x_*, λ_*).*

Then, if (x_k, λ_k) is sufficiently close to (x_*, λ_*), there is a solution of the SQP subproblem whose active set $\mathcal{A}_k$ is the same as the active set $\mathcal{A}(x_*)$.

- Consider solving the inequality constrained quadratic problem

$$\begin{aligned}\min_{d}\ & f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T \nabla_{xx}^2 L(x_k, \lambda_k) d \\ \text{s.t. } & c^i(x_k) + \nabla c^i(x_k)^T d = 0,\ \ i \in \mathcal{E} \\ & c^i(x_k) + \nabla c^i(x_k)^T d \leq 0,\ \ i \in \mathcal{I}\end{aligned}$$

- Estimate $\mathcal{A}_k$, then solve the equality constrained quadratic problem

$$\begin{aligned}\min_{d}\ & f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T \nabla_{xx}^2 L(x_k, \lambda_k) d \\ \text{s.t. } & c^i(x_k) + \nabla c^i(x_k)^T d = 0,\ \ i \in \mathcal{E} \cup \mathcal{A}_k.\end{aligned}$$

We can use a ℓ_1 merit function for judging progress, as before

$$\phi(x;\nu) := f(x) + \nu\|c^{\mathcal{E}}(x)\|_1 + \nu\|\max\{c^{\mathcal{I}}(x), 0\}\|_1$$

Then the directional derivative of the ℓ_1 merit function equals

$$D\phi(d_k; x_k, \nu) = \nabla f(x_k)^T d_k + \nu \sum_{i\in\mathcal{E}} |\nabla c^i(x_k)^T d| + \nu \sum_{i\in\mathcal{I}} \max\{\nabla c^i(x_k)^T d, 0\}$$

For a sufficiently large ν, we can guarantee

$$D\phi(d_k; x_k, \nu) << 0.$$

An alternative is to relax the constraints:

$$c(x_k) + \nabla c(x_k)^T d = r - s, \;\; (r, s) \geq 0$$

The SQP subproblem can be reformulated as

$$\begin{aligned} \min_d \; & f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T H_k d + \nu \sum_{i \in \mathcal{E}} (r^i + s^i) + \nu \sum_{i \in \mathcal{I}} t^i \\ \text{s.t. } & c^{\mathcal{E}}(x_k) + \nabla c^{\mathcal{E}}(x_k)^T d = r - s, \quad (r, s) \geq 0 \\ & c^{\mathcal{I}}(x_k) + \nabla c^{\mathcal{I}}(x_k)^T d = t, \quad t \geq 0 \end{aligned}$$