Quadratic programs and Active set methods

Eddie Wadbro

Introduction to PDE Constrained Optimization, 2016

February 15–16, 2016

(i) Equality-constrained QP's

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}x^{\mathsf{T}}Qx - c^{\mathsf{T}}x \quad \text{subject to}$$
$$a_i^{\mathsf{T}}x = b_i \qquad i = 1, \ldots, m$$

or

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}x^{\mathsf{T}}Qx - c^{\mathsf{T}}x \text{ subject to}$$
$$Ax = b$$

where Q is symmetric, $m < n$ and $A = \begin{pmatrix} a_1^{\mathsf{T}} \\ \vdots \\ a_m^{\mathsf{T}} \end{pmatrix}$. The Lagrangian is

$$\begin{aligned}\mathcal{L}(x,\lambda) &= \frac{1}{2}x^{\mathsf{T}}Qx - c^{\mathsf{T}}x - \sum_{i=1}^{m}\lambda_i(a_i^{\mathsf{T}}x - b_i)\\ &= \frac{1}{2}x^{\mathsf{T}}Qx - c^{\mathsf{T}}x - \lambda^{\mathsf{T}}(Ax - b)\end{aligned}$$

KKT-points for equality-constrained QP's

KKT-point $\nabla_x\mathcal{L}(x^*,\lambda^*) = 0$ yields the linear system

$$\begin{aligned}Qx^* - A^{\mathsf{T}}\lambda^* &= c\\ Ax^* &= b\end{aligned}$$

or

$$\begin{pmatrix} Q & -A^{\mathsf{T}} \\ A & 0 \end{pmatrix}\begin{pmatrix} x^* \\ \lambda^* \end{pmatrix} = \begin{pmatrix} c \\ b \end{pmatrix}$$

where

$$\mathcal{K} = \begin{pmatrix} Q & -A^{\mathsf{T}} \\ A & 0 \end{pmatrix}$$

is called a **KKT matrix**. Write the constraint $-Ax = -b$, substitute $B = -A$. Then, the KKT matrix becomes symmetric:

$$\begin{pmatrix} Q & B^{\mathsf{T}} \\ B & 0 \end{pmatrix}\begin{pmatrix} x^* \\ \lambda^* \end{pmatrix} = \begin{pmatrix} c \\ -b \end{pmatrix}$$

Optimality for equality-constrained QP's

Assumptions:

- A has linearly independent rows.
- Q is positive definite in the null space of A. (That is, $x^T Qx > 0$ for all $x \neq 0$ such that $Ax = 0$).

Theorem

Matrix $\mathcal{K}$ is nonsingular.

Theorem

The solution x^* of the KKT system is the unique global solution of the equality-constrained QP.

- The equality-constrained QP is a convex problem under the above assumptions.
- Note: matrix $\mathcal{K}$ is indefinite.
- Thus, solving a convex equality-constraint QP is "easy"
- Equivalent to solving a linear system (the KKT system)

(ii) Inequality-constrained QP's

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^{\mathsf{T}} Qx - c^{\mathsf{T}} x \text{ subject to}$$
$$Ax \geq b$$

Lagrangian: $\mathcal{L}(x, \lambda) = \frac{1}{2} x^{\mathsf{T}} Qx - c^{\mathsf{T}} x - \lambda^{\mathsf{T}}(Ax - b)$

KKT conditions:

$$\begin{aligned} Qx^* - A^{\mathsf{T}} \lambda^* &= c \\ \lambda^* &\geq 0 \\ Ax^* &\geq b \\ \lambda_i^*(a_i^{\mathsf{T}} x^* - b_i) &= 0 \quad i = 1, \ldots, m \end{aligned}$$

Define the **active set**

$$\mathcal{A} = \left\{ i \mid a_i^{\mathsf{T}} x^* = b_i \right\}$$

Note that $\lambda_i^* = 0$ for $i \notin \mathcal{A}$

Optimality conditions for inequality-constrained QP's

- We may delete all inactive inequality constraints and corresponding zero Lagrange multipliers
- Let $\bar{A}$ be A with all rows for $i \notin \mathcal{A}$ deleted
- Let $\bar{b}$ be b with all rows for $i \notin \mathcal{A}$ deleted
- Let $\bar{\lambda}^*$ be λ^* with all components for $i \notin \mathcal{A}$ deleted.

Then the KKT conditions simplify to

$$\begin{aligned} Qx^* - \bar{A}^{\mathsf{T}} \bar{\lambda}^* &= c \\ \bar{A}x^* &= \bar{b} \end{aligned}$$

i. e. the KKT conditions for a QP with **equality** constraints

Note: The above form assume that $\mathcal{A}$ is known (which it generally **not** is!)

Background to active set method for inequality constrained QP

- An active-set method generates feasible points
- Assume that we know a feasible point x_k (can be obtained via a linear problem)
- Define a **working set** with constraints active at the current iterate
$$\mathcal{W}_k = \left\{ x \mid a_i^{\mathsf{T}} x_k = b_i \right\}$$
- Guess that the constraints active at x are active at x^* too. That is keep (temporarily) the constraints in $\mathcal{W}$ active and solve
$$\begin{aligned} \min \quad & \frac{1}{2}(x_k + p)^{\mathsf{T}} Q (x_k + p) + c^{\mathsf{T}}(x_k + p) \\ \text{subject to} \quad & A_{\mathcal{W}} p = 0, \end{aligned} \qquad \text{(EQP)}$$
where $A_{\mathcal{W}}$ equals A with all rows for $i \notin \mathcal{W}_k$ deleted

Active set for QP

Problem (EQP) has, from above, optimal solution p^* and associated Lagrange multiplier vector λ^* given by
$$\begin{pmatrix} Q & -A_{\mathcal{W}}^{\mathsf{T}} \\ -A_{\mathcal{W}} & 0 \end{pmatrix} \begin{pmatrix} p^* \\ \lambda^* \end{pmatrix} = - \begin{pmatrix} Qx_k - c \\ 0 \end{pmatrix}.$$
Optimal x associated with (EQP) is given by $x^* = x_k + p$.

When solving (EQP) we have ignored two things

1. All inactive constraints, that is, we must require $a_i^{\mathsf{T}} x \geq b_i$ for $i \notin \mathcal{W}$.
2. The constraints are inequalities, we have required $A_{\mathcal{W}} p = 0$ instead of $A_{\mathcal{W}} p \geq 0$.

How are these requirements included?

Inclusion of new constraints

We have started in x_k and computed a search direction p^*.

If $A(x_k + p^*) \geq b$, then $x_k + p^*$ satisfies all constraints.

Otherwise, we can compute the maximum step-length α so that $A(x_k + \alpha p^*) \geq b$ holds. Thus, we compute
$$\alpha = \min_{i \mid a_i^{\mathsf{T}} p^* < 0} \frac{a_i^{\mathsf{T}} x_k - b_i}{-a_i^{\mathsf{T}} p^*},$$
define $x_{k+1} = x_k + \alpha p^*$, and set $\mathcal{W} = \mathcal{W} \cup \{l\}$; where $a_l^{\mathsf{T}} x_{k+1} = b_l$.

Removal of constraints

The point $x_k + p^*$ is of interest when $A(x_k + p^*) \geq b$.

When solving (EQP) we obtain p^* and λ^*.

Two cases:

1. $\lambda^* \geq 0$. Then $x^* = x_k + p^*$ is the optimal solution to
$$\begin{aligned} \min \quad & \frac{1}{2}x^T Q x - c^T x \\ \text{subject to} \quad & A_{\mathcal{W}} x \geq b_{\mathcal{W}}, \end{aligned}$$
and hence an optimal solution to the original inequality constrained QP!
2. $\lambda_i^* < 0$ for some i. Let $x_{k+1} = x_k + p^*$ and set $\mathcal{W} = \mathcal{W} \setminus \{i\}$.

Sequential Quadratic Programming (SQP)

- One of the most efficient methods for nonlinear programming.
- Recommended as a general purpose method for small to medium scale problems.
- (e.g. `fmincon` medium scale is a SQP)

Consider the problem
$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & g(x) = 0, \end{aligned} \tag{NLP}$$
where $f : \mathbb{R}^n \mapsto \mathbb{R}$, and $g : \mathbb{R}^n \mapsto \mathbb{R}^m$, that is, $g = \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{pmatrix}$.

Lagrangian and KKT-conditions

The Lagrangian of NLP is
$$\mathcal{L}(x, \lambda) = f(x) - \lambda^T g(x) = f(x) - \sum_{i=1}^{m} \lambda_i g_i(x).$$

KKT-conditions
$$\begin{cases} \nabla_x f(x^*) - \sum \lambda_i \nabla_x g_i(x^*) = 0 \\ -g_i(x^*) = 0 \end{cases} \iff \begin{cases} \nabla_x \mathcal{L}(x, \lambda) = 0 \\ \nabla_\lambda \mathcal{L}(x, \lambda) = 0 \end{cases}$$
Thus solving the KKT-system is equivalent to solving
$$\nabla \mathcal{L}(x, \lambda) = 0, \tag{OC}$$
where $\nabla = (\nabla_x, \nabla_\lambda)^T$.

Recall: Newtons method

For unconstrained optimization a necessary condition for minimum is $\nabla f(x^*) = 0$,

$$\nabla_x f(x_k + p_k) = [Taylor] \ldots \approx \nabla_x f(x_k) + \nabla_x^2 f(x_k) p_k = 0$$
$$\iff \quad \nabla_x^2 f(x_k) p_k = -\nabla_x f(x_k) \quad \text{(BN)}$$

i.e. Newton's metod

SQP

Do the same with (OC)

$$\nabla \mathcal{L}(x_k + p_k, \lambda_k + \mu_k) = [Taylor] \ldots$$
$$\approx \nabla \mathcal{L}(x_k, \lambda_k) + \nabla^2 \mathcal{L}(x_k, \lambda_k) \begin{pmatrix} p_k \\ \mu_k \end{pmatrix} = 0$$

Thus to find our steps p_k and μ_k, we solve the system

$$\nabla^2 \mathcal{L}(x_k, \lambda_k) \begin{pmatrix} p_k \\ \mu_k \end{pmatrix} = -\nabla \mathcal{L}(x_k, \lambda_k), \qquad \text{(SQP)}$$

where

$$\nabla^2 \mathcal{L}(x_k, \lambda_k) = \begin{pmatrix} \nabla_x^2 \mathcal{L}(x_k, \lambda_k) & -\nabla_x g(x_k) \\ -\nabla_x g(x_k)^{\mathsf{T}} & 0 \end{pmatrix}$$

with

$$\nabla_x g = \begin{pmatrix} \nabla_x g_1 & \nabla_x g_2 & \ldots & \nabla_x g_m \end{pmatrix}.$$

(SQP) is the basic SQP method (just as (BN) is the basic Newton).

Properties of SQP

Similar properties as Newton's method: Quadratic convergence rate if

(i) $\nabla^2 \mathcal{L}(x^*, \lambda^*)$ is nonsingular
(ii) started close enough.

$$\text{(SQP)} \iff \begin{pmatrix} \nabla_x^2 \mathcal{L}(x_k, \lambda_k) & -\nabla_x g(x_k) \\ -\nabla_x g(x_k)^{\mathsf{T}} & 0 \end{pmatrix} \begin{pmatrix} p_k \\ \mu_k \end{pmatrix} = -\nabla \mathcal{L}(x_k, \lambda_k),$$

optimality system to the QP

$$\min_p \quad \frac{1}{2} p^{\mathsf{T}} \nabla_x^2 \mathcal{L}(x_k, \lambda_k) p + p^{\mathsf{T}} \nabla_x \mathcal{L}(x_k, \lambda_k)$$
$$\text{s.t.} \quad \underbrace{\nabla_x g(x_k)^{\mathsf{T}} p + g(x_k)}_{\approx g(x_k + p)} = 0,$$

The minimization of a quadratic approximation of the Lagrangian subject to a linearization of the constraints (therefore the name SQP).

Modifications

Basic SQP benefits fro msimilar modifications as basic Newton

(i) Newton direction is a descent direction for unconstrained optimization if Hessian (or the approximation) is PD.
The QP has a unique min if $\nabla_x^2 \mathcal{L}$ is PD in the nullspace of $(\nabla_x g)^\mathsf{T}$.
Need to ensure this...

(ii) Unconstrained optimization: line search to ensure

$$f(x_k + \alpha_k p_k) \leq f(x_k) + \mu \alpha_k \underbrace{p_k^\mathsf{T} \nabla_x f(x_k)}_{\leq 0}.$$

For SQP, monitor prograss through a **merit function** φ, for example

$$\varphi(x) = f(x) + \frac{1}{2\mu} \sum_{i=1}^{m} g_i(x)^2$$

(quadratic penalty).

Inequality constraints

Inequality constraints can be handled by linearizing them and then use an active set strategy in the QP subproblem.