Lecture 14

Contents[1]

1. A quick review of Quadratic programming with equality constraints.
2. Active set method for Quadratic programming with inequality constraints, and an example illustrating the same.
3. Quadratic penalty method with equality constraints.

Quadratic programming with Equality Constraints

The objective is to,

$$\min_{x \in \mathcal{R}^n} \frac{1}{2}x^T G x + x^T d, \quad s.t \quad Ax = b \tag{1}$$

Then by applying Lagrangian on Eq. 1 we have,

$$L(x^*, \lambda^*) = \frac{1}{2}x^{*T} G x^* + x^{*T} d - \lambda^{*T}(Ax^* - b) \tag{2}$$

By applying KKT conditions on Eq. 2, one can arrive at following linear system of equations as shown below.

$$\begin{bmatrix} G & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} -d \\ b \end{bmatrix} \tag{3}$$

The system (3) can be re-written in a form that is useful for computation by expressing x^* as $x + p$ where x is some estimate of the solution and p is the desired step. Then the system (3) gets modified as,

$$\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} -p \\ \lambda^* \end{bmatrix} = \begin{bmatrix} g \\ c \end{bmatrix} \tag{4}$$

where,

$$\begin{aligned} c &= Ax - b \\ g &= d + Gx \\ p &= x^* - x \end{aligned}$$

By eliminating the block of equations involving p we have,

$$\left(AG^{-1}A^T\right)\lambda^* = \left(AG^{-1}g - c\right) \tag{5}$$

Thus, we can solve the system (5) for λ^* and then we can obtain p by solving system (4) and finally the optima is given by $x^* = x + p$.

[1] This scribe has been made with reference to the book "Numerical Optimization" by Nocedal & Wright.

Quadratic programming with Inequality Constraints

Note that when inequality constraints are present in the given optimization problem, we usually don't have the prior knowledge of the active set $\mathcal{A}(x^*)$.

Working Set

Working set ($\mathcal{W}$) is a subset of constraints. It consists of all the equality constraints $i \in \mathcal{E}$ together with some - but not necessary all - of the active inequality constraints. An important requirement we impose on $\mathcal{W}$ is that the gradients of the constraints in the working set be linearly independent, even when the full set of active constraints at that point has linearly dependent gradients.

Active set method

Given an iterate x_k and the working set $\mathcal{W}_k$, we first check whether x_k minimizes the quadratic function in the subspace defined by the working set. If not, we compute a step p by solving an equality constrained QP subproblem (as discussed earlier) in which the constraints corresponding to $\mathcal{W}_k$ are regarded as equalities and all other constraints are temporarily discarded. Putting it all together, our objective is,

$$\min_{p \in \mathcal{R}^n} \frac{1}{2}p^T G p + g_k^T p + y, \quad s.t \quad a_i^T p = 0 \quad \forall i \in \mathcal{W}_k \tag{6}$$

where,

$$y = \frac{1}{2}x_k^T G x_k + x_k^T d, \quad g_k = d + Gx_k, \quad p = x - x_k$$

Let us denote the solution of this subproblem by p_k. Note that for each $i \in \mathcal{W}_k$, the term $a_i^T x$ does not change as we move along p_k, since we have $a_i^T(x_k + p_k) = a_i^T x_k = b_i$. If $x_k + p_k$ is feasible with respect to all the constraints, we set $x_{k+1} = x_k + p_k$, otherwise we set $x_{k+1} = x_k + \alpha_k p_k$. Note that $a_i^T p_k \geq 0 \ \forall i \in \mathcal{W}_k$.

We can derive an explicit definition of α_k by considering the constraints $i \notin \mathcal{W}_k$. Whenever $a_i^T p_k < 0$ for some $i \notin \mathcal{W}_k$, we have $a_i^T(x_k + \alpha_k p_k) \geq b_i$ only if,

$$\alpha_k \leq \frac{b_i - a_i^T x_k}{a_i^T p_k} \tag{7}$$

Since we want α_k to be as large as possible in $[0, 1)$ subject to retaining feasibility, we have,

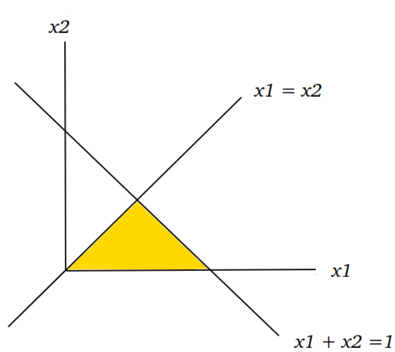
$$\alpha_k = \min\left(1, \min_{\substack{i \notin \mathcal{W}_k \\ a_i^T p_k < 0}} \frac{b_i - a_i^T x_k}{a_i^T p_k}\right) \tag{8}$$

Algorithm 1 Active set method for Convex QP

Input : A feasible starting point x_0;
Output : x^*;
$\mathcal{W}_0 \leftarrow$ *A subset of the active constraints at* x_0 ;
for $k = 0, 1, 2, 3, \ldots$ **do**
    solve Eq. 6 for p_k;
    if $p_k = 0$ **then**
        compute the lagrange multipliers λ_i;
        if $\lambda_i \geq 0 \ \forall i \in \mathcal{W}_k \cap \mathcal{I}$ **then**
            Stop *and* $x^* \leftarrow x_k$;
        else
            $j \leftarrow \min_{j \in \mathcal{W}_k \cap \mathcal{I}} \lambda_j$;
            $x_{k+1} \leftarrow x_k$;
            $\mathcal{W}_{k+1} \leftarrow \mathcal{W}_k \setminus \{j\}$;
        end if
    else
        compute α_k *from Eq. 8*;
        $x_{k+1} \leftarrow x_k + \alpha_k p_k$;
        if *there are blocking constraints* **then**
            $\mathcal{W}_{k+1} \leftarrow \mathcal{W}_k +$ *one of the blocking constraint*;
        else
            $\mathcal{W}_{k+1} \leftarrow \mathcal{W}_k$;
        end if
    end if
end for

Example

$$\min_{\mathbf{x}} \frac{1}{2}\left(\left(x_1 - 3\right)^2 + \left(x_2 - 2\right)^2\right) \quad s.t \quad x_1 - x_2 \geq 0, \ -x_1 - x_2 \geq -1, \ x_2 \geq 0$$



$$G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, d = \begin{bmatrix} -3 \\ -2 \end{bmatrix}$$

Let,

$$\mathbf{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathcal{W}_0 = \{1, 3\} \implies A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}, b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

where,

$$x_1 - x_2 = 0 \rightarrow \{1\}$$
$$x_2 = 0 \rightarrow \{3\}$$

Iteration 0

Thus, for given $\mathcal{W}_0$, the feasible set is $\{(0, 0)\}$. This implies that $\mathbf{p} = [0, 0]^T$. Now, we have

$$g_0 = Gx_0 + d = \begin{bmatrix} -3 \\ -2 \end{bmatrix}$$

$$c_0 = Ax_0 - b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Remember that,

$$\left(AG^{-1}A^T\right)\lambda^* = \left(AG^{-1}g_0 - c_0\right)$$

As c_0 is a zero vector we have,

$$A^T\lambda^* = g_0$$

On solving for λ^*,

$$\begin{bmatrix} \lambda_1 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} -3 \\ -5 \end{bmatrix}$$

Iteration 1

Thus,

$$\mathbf{x}_1 = \mathbf{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies g_1 = g_0, \;\; \mathcal{W}_1 = \mathcal{W}_0 \setminus \{3\} = \{1\} \implies A = \begin{bmatrix} 1 & -1 \end{bmatrix}, b = \begin{bmatrix} 0 \end{bmatrix}, c_1 = \begin{bmatrix} 0 \end{bmatrix}$$

For given $\mathcal{W}_1$, the feasible set is $\{\mathbf{x} \mid x_1 = x_2\}$. Now, we have the KKT system as shown below.

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} -p_1 \\ -p_2 \\ \lambda_1 \end{bmatrix} = \begin{bmatrix} -3 \\ 2 \\ 0 \end{bmatrix}$$

On solving above system,

$$\begin{bmatrix} p_1 \\ p_2 \\ \lambda_1 \end{bmatrix} = \begin{bmatrix} 5/2 \\ 5/2 \\ -1/2 \end{bmatrix}$$

Now, if we take a complete step we will violate $\{2\}$ and thus we need to find α_1.

$$\alpha_1 = \min\left(1, \min_{\substack{i \notin \mathcal{W}_1 \\ a_i^T p_k < 0}} \frac{b_i - a_i^T x_k}{a_i^T p_k}\right) = \frac{1}{5}$$

Iteration 2

Thus,

$$\mathbf{x}_2 = \mathbf{x}_1 + \alpha_1 \mathbf{p}_1 = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}, g_2 = \begin{bmatrix} -5/2 \\ -3/2 \end{bmatrix}, \mathcal{W}_2 = \{2\} \implies A = \begin{bmatrix} -1 & -1 \end{bmatrix}, b = \begin{bmatrix} -1 \end{bmatrix}, c_2 = \begin{bmatrix} 0 \end{bmatrix}$$

For given $\mathcal{W}_2$, the feasible set is $\{\mathbf{x} \mid x_1 + x_2 = 1\}$. Now, we have the KKT system as shown below.

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 0 \end{bmatrix} \begin{bmatrix} -p_1 \\ -p_2 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} -5/2 \\ -3/2 \\ 0 \end{bmatrix}$$

On solving above system,

$$\begin{bmatrix} p_1 \\ p_2 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 1/2 \\ -1/2 \\ 2 \end{bmatrix}$$

Iteration 3

Thus,

$$\mathbf{x}_3 = \mathbf{x}_2 + \mathbf{p}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, g_3 = \begin{bmatrix} -2 \\ -2 \end{bmatrix}, \mathcal{W}_3 \leftarrow \mathcal{W}_2 \implies A = \begin{bmatrix} -1 & -1 \end{bmatrix}, b = \begin{bmatrix} -1 \end{bmatrix}, c_3 = \begin{bmatrix} 0 \end{bmatrix}$$

For given $\mathcal{W}_3$, the feasible set is $\{\mathbf{x} \mid x_1 + x_2 = 1\}$. Now, we have the KKT system as shown below.

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 0 \end{bmatrix} \begin{bmatrix} -p_1 \\ -p_2 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \\ 0 \end{bmatrix}$$

On solving above system,

$$\begin{bmatrix} p_1 \\ p_2 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

As $\mathbf{p}_3 = 0$ and $\lambda_2 > 0$, we can conclude that $\mathbf{x}_3 = [1, 0]^T$ is the optima for the given constrained Quadratic problem.

Quadratic Penalty methods with Equality Constraints

One fundamental approach to constrained optimization problem is to replace the original problem by a penalty function that consists of,

1. The original objective of the constrained optimization problem, plus
2. One additional term for each constraint, which is positive when the current point $\mathbf{x}$ that violates that constraint and zero otherwise.

The simplest penalty function of this type is the quadratic penalty function, in which the penalty terms are the squares of the constraint violations. Consider,

$$\min_{\mathbf{x}} f(\mathbf{x}), \quad s.t \quad c_i(\mathbf{x}) = 0, \forall i \in \mathcal{E} \tag{9}$$

The quadratic penalty function for Eq. 9 is of the form,

$$\min_{\mathbf{x}} Q(\mathbf{x}; \mu) = f(\mathbf{x}) + \frac{1}{2\mu} \sum_{i \in \mathcal{E}} c_i^2(\mathbf{x}) \tag{10}$$

where $\mu > 0$, is the penalty parameter.

Example

$$\min_{\mathbf{x} \in \mathcal{R}^2} \quad x_1 + x_2 \quad s.t \quad x_1^2 + x_2^2 - 2 = 0$$

The contour plots of $Q(\mathbf{x}; \mu)$ for $\mu = 1$ and $\mu = 100$ are shown in Fig.1 and Fig.2 respectively. We observe that for $\mu = 100$ the minimizer (white *) is more close to actual optima (yellow star).

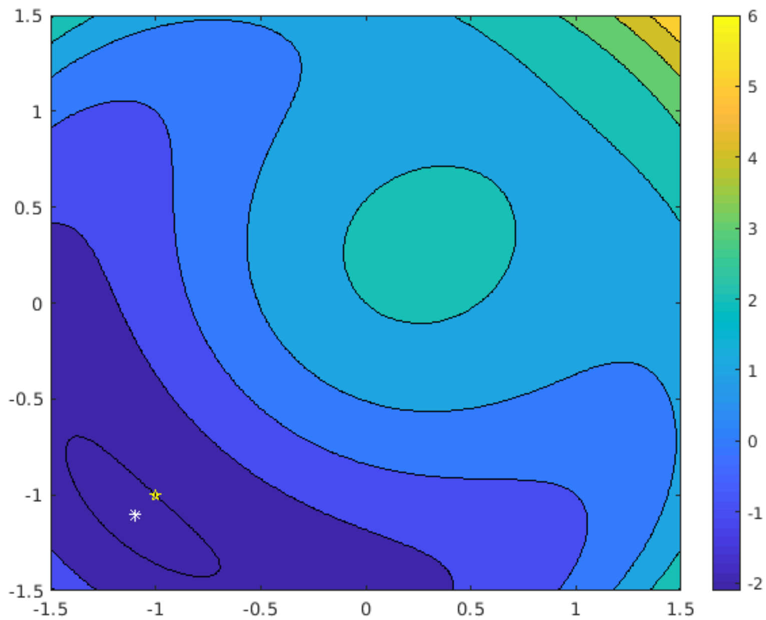


Figure 1: Contours of $Q(\mathbf{x};\mu)$ for $\mu = 1$

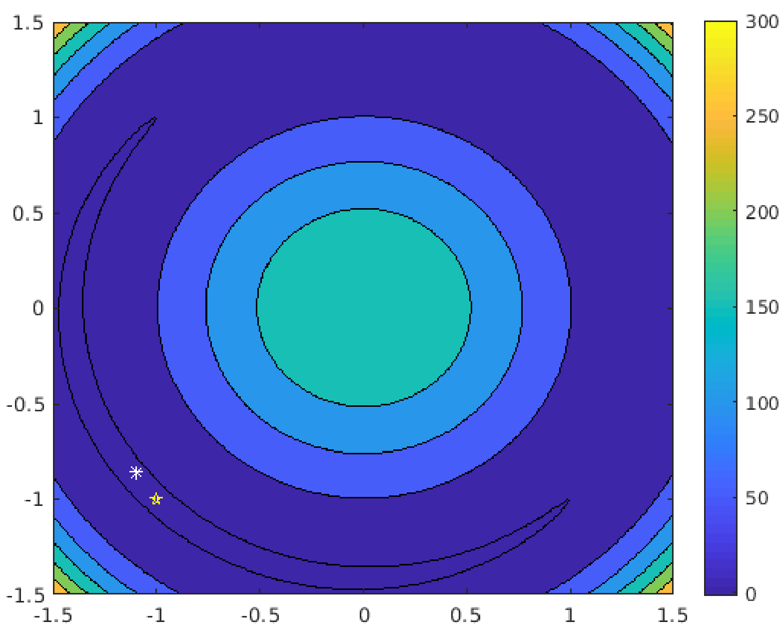


Figure 2: Contours of $Q(\mathbf{x};\mu)$ for $\mu = 100$