

CMO: Active Set Method

Eklavya Sharma

The active set method is a way of solving optimization problems with a convex quadratic objective and linear constraints. In this method, we guess the active set of constraints and then solve the resulting problem which has only equality constraints. We make multiple such guesses and each guess helps refine the next guess.

1 Equality constraints

We'll first figure out how to solve the optimization problem which has only equality constraints:

$$\min_x f(x) \text{ where } Ax = b$$

Here $f(x) = \frac{1}{2}x^T Qx - h^T x$, where Q is symmetric and positive definite. The i^{th} row of A is a_i^T . $x \in \mathbb{R}^d$ and $b \in \mathbb{R}^m$.

We'll find a KKT point for this problem. Since the objective is strictly convex and the constraints are linear, the KKT point gives us the unique global minimum of this problem.

$$L(x, \mu) = f(x) - \mu^T (Ax - b)$$

By stationarity, we get

$$\nabla_f(x) - A^T \mu = 0 \implies Qx - A^T \mu = h$$

By primal feasibility, we get

$$Ax = b$$

These 2 conditions can be written as a system of linear equations. Solving this system will give us the KKT point (x, λ) .

$$\begin{bmatrix} Q & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \mu \end{bmatrix} = \begin{bmatrix} h \\ b \end{bmatrix}$$

This system of equations is guaranteed to have a solution, since (x, λ) is a KKT point iff it satisfies these conditions and the first-order necessary conditions for local minimum imply that a KKT point must exist (regularity is ensured because the constraints are linear).

2 Inequality constraints

We'll now try to solve this optimization problem:

$$\min_x f(x) \text{ where } Ax \geq b$$

Define argminset as

$$\text{argminset } g(x) = \{\hat{x} : g(\hat{x}) = \min_x g(x)\}$$

The subroutine $\text{eqsolve}(Q, h, A, b)$ returns (x^*, μ^*) such that

$$\begin{bmatrix} Q & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \mu^* \end{bmatrix} = \begin{bmatrix} h \\ b \end{bmatrix}$$

This means that

$$x^* = \underset{\substack{x \\ Ax=b}}{\text{argmin}} \frac{1}{2} x^T Q x - h^T x$$

The subroutine $\text{blocking-constraints}$ is defined as

$$\text{blocking-constraints}(A, x, u) = \underset{a_i^T u < 0}{\text{argminset}} \frac{a_i^T x - b_i}{-a_i^T u}$$

Let A_B denote the matrix obtained by taking all rows of A whose indices lie in B .

Algorithm 1 active-set-method($Q, h, A, b, x^{(0)}$): x_0 is the initial feasible point

```

1:  $B^{(0)} = \{i : a_i^T x^{(0)} = b_i\}$ 
2:  $t = 0$ 
3: while true do
4:   while true do
5:      $(u^{(t)}, \mu^{(t)}) = \text{eqsolve}(Q, h - Qx^{(t)}, A_{B^{(t)}}, 0)$ 
6:     if  $u^{(t)} == 0$  then
7:       break
8:     end if
9:      $I^{(t)} = \text{blocking-constraints}(A, x^{(t)}, u^{(t)})$ 
10:     $\alpha^{(t)} = \left(1 \text{ if } I^{(t)} == \{\} \text{ else } \min\left(1, \frac{a_i^T x^{(t)} - b_i}{-a_i^T u^{(t)}}\right)\right)$ , where  $i \in I^{(t)}$ .
11:     $x^{(t+1)} = x^{(t)} + \alpha^{(t)} u^{(t)}$ 
12:     $B^{(t+1)} = B^{(t)} \cup (I^{(t)} \text{ if } \alpha^{(t)} < 1 \text{ else } \{\})$ 
13:     $t = t + 1$ 
14:  end while
15:  Find  $l^{(t)} \in B^{(t)}$  such that  $\mu_{l^{(t)}}^{(t)} < 0$ .
16:  if  $l^{(t)} == \text{null}$  then
17:     $\lambda_i = \begin{cases} \mu_i^{(t)} & i \in B^{(t)} \\ 0 & i \notin B^{(t)} \end{cases}$ 
18:    return  $(x^{(t)}, \lambda)$ 
19:  else
20:     $B^{(t+1)} = B^{(t)} - \{l^{(t)}\}$ 
21:     $x^{(t+1)} = x^{(t)}$ 
22:     $t = t + 1$ 
23:  end if
24: end while

```

Lemma 1 (Invariants). *Let $\mathcal{A}(x) = \{i : a_i^T x = b_i\}$. The following are true for all t :*

- $x^{(t)}$ is a feasible solution.
- $B^{(t)} \subseteq \mathcal{A}(x^{(t)})$.

Proof by induction.

Base case: $x^{(0)}$ is given to be feasible. $B^{(0)} = \mathcal{A}(x^{(0)})$ by line 1. Therefore, the hypothesis holds for $t = 0$.

Inductive step: Assume the hypothesis holds at t .

Case 1: $u^{(t)} \neq 0$

If $u^{(t)} \neq 0$, then $x^{(t+1)} = x^{(t)} + \alpha^{(t)} u^{(t)}$. Let $J = \{i : a_i^T u^{(t)} < 0\}$. Then $\forall i \in J$,

$$\alpha^{(t)} \leq \frac{a_i^T x^{(t)} - b_i}{-a_i^T u^{(t)}} \implies b_i \leq a_i^T (x^{(t)} + \alpha^{(t)} u^{(t)}) = a_i^T x^{(t+1)}$$

When $i \notin J$, then $a_i^T u^{(t)} \geq 0$. Also, $a_i^T x^{(t)} \geq b_i$, by inductive hypothesis. Therefore,

$$a_i^T x^{(t+1)} = a_i^T x^{(t)} + \alpha^{(t)} a_i^T u^{(t)} \geq b_i$$

Therefore, $x^{(t+1)}$ is feasible.

$$i \in B^{(t)} \implies a_i^T x^{(t)} = b_i \quad (B^{(t)} \subseteq \mathcal{A}(x^{(t)}) \text{ by inductive hypothesis})$$

$$(u^{(t)}, \mu^{(t)}) = \text{eqsolve}(Q, h - Qx^{(t)}, A_{B^{(t)}}, 0) \implies \forall i \in B^{(t)}, a_i^T u^{(t)} = 0$$

Therefore, $\forall i \in B^{(t)}$,

$$a_i^T x^{(t+1)} = a_i^T x^{(t)} + \alpha^{(t)} a_i^T u^{(t)} = b_i$$

Therefore, $B^{(t)} \subseteq \mathcal{A}(x^{(t+1)})$.

If $\alpha^{(t)} \geq 1$ or $I^{(t)} = \{\}$, then $B^{(t+1)} = B^{(t)} \subseteq \mathcal{A}(x^{(t+1)})$. Suppose $\alpha^{(t)} < 1$ and $I^{(t)} \neq \{\}$. Then $B^{(t+1)} = B^{(t)} \cup I^{(t)}$ and $\forall i \in I^{(t)}$,

$$\alpha^{(t)} = \frac{a_i^T x^{(t)} - b_i}{-a_i^T u^{(t)}} \implies b_i = a_i^T (x^{(t)} + \alpha^{(t)} u^{(t)}) = a_i^T x^{(t+1)}$$

Therefore, $I^{(t)} \subseteq \mathcal{A}(x^{(t+1)})$ which implies that $B^{(t+1)} \subseteq \mathcal{A}(x^{(t+1)})$.

Case 2: $u^{(t)} = 0$

By inductive hypothesis, $x^{(t)}$ is feasible. Therefore, $x^{(t+1)} = x^{(t)}$ is also feasible. By inductive hypothesis, $B^{(t)} \subseteq \mathcal{A}(x^{(t)})$. Therefore, $B^{(t+1)} \subset B^{(t)} \subseteq \mathcal{A}(x^{(t+1)})$.

Therefore, the inductive hypothesis holds at $t + 1$. By mathematical induction, the inductive hypothesis holds for all $t \geq 0$. \square

Theorem 2 (Correctness). *If active-set-method returns $(x^{(t)}, \lambda)$, then $x^{(t)}$ is a global minimum.*

Proof. We'll prove that $(x^{(t)}, \lambda)$ is a KKT point. This is a sufficient condition for minimum since this is a convex optimization problem.

The algorithm enters the conditional block on line 16 iff $u^{(t)} = 0$ and $\mu_i^{(t)} \geq 0$ for all $i \in B^{(t)}$.

$$\begin{aligned} (0, \mu) &= \text{eqsolve}(Q, h - Qx^{(t)}, A_{B^{(t)}}, 0) \\ \implies -A_{B^{(t)}}^T \mu &= h - Qx^{(t)} \\ \implies \nabla_f(x^{(t)}) &= A_{B^{(t)}}^T \mu = \sum_{i \in B^{(t)}} \mu_i a_i = \sum_{i=1}^m \lambda_i a_i \end{aligned}$$

This proves stationarity for $(x^{(t)}, \lambda)$.

Primal feasibility follows from lemma 1.

Dual feasibility follows from the fact that $\lambda_i = \mu_i^{(t)} \geq 0$ for all $i \in B^{(t)}$ and $\lambda_i = 0$ for all $i \notin B^{(t)}$.

When $i \notin B^{(t)}$, then $\lambda_i = 0$. When $i \in B^{(t)}$, then $a_i^T x^{(t)} = b_i$ (since $B^{(t)} \subseteq \mathcal{A}(x^{(t)})$ by lemma 1). This ensures complementary slackness.

Therefore, $(x^{(t)}, \lambda)$ is a KKT point. \square

Lemma 3. $B^{(t)} = B^{(t+1)} \implies u^{(t+1)} = 0$

Proof.

$$\begin{aligned}
(I^{(t)} = \{\}) &\implies \alpha^{(t)} = 1 \\
&\implies (\alpha^{(t)} < 1 \implies I^{(t)} \neq \{\} \implies B^{(t+1)} \neq B^{(t)}) \\
&\implies (B^{(t+1)} = B^{(t)} \implies \alpha^{(t)} = 1) \\
&\implies (B^{(t+1)} = B^{(t)} \implies x^{(t+1)} = x^{(t)} + u^{(t)})
\end{aligned}$$

By lemma 1, $x^{(t)} + u^{(t)}$ is feasible.

$$\begin{aligned}
(u^{(t)}, \mu^{(t)}) &= \text{eqsolve}(Q, h - Qx^{(t)}, A_{B^{(t)}}, 0) \\
&\implies u^{(t)} = \underset{A_{B^{(t)}} u = 0}{\operatorname{argmin}} f(x^{(t)} + u) \\
&\implies x^{(t+1)} = x^{(t)} + u^{(t)} = \underset{\forall i \in B^{(t)}, a_i^T x = b_i}{\operatorname{argmin}} f(x) \\
&\implies x^{(t+1)} = \underset{\forall i \in B^{(t+1)}, a_i^T x = b_i}{\operatorname{argmin}} f(x) \quad (B^{(t+1)} = B^{(t)})
\end{aligned}$$

$$\begin{aligned}
(u^{(t+1)}, \mu^{(t+1)}) &= \text{eqsolve}(Q, h - Qx^{(t+1)}, A_{B^{(t+1)}} , 0) \\
&\implies u^{(t+1)} = \underset{A_{B^{(t+1)}} u = 0}{\operatorname{argmin}} f(x^{(t+1)} + u) \\
&\implies x^{(t+1)} + u^{(t+1)} = \underset{\forall i \in B^{(t+1)}, a_i^T x = b_i}{\operatorname{argmin}} f(x)
\end{aligned}$$

Therefore, $x^{(t+1)} = x^{(t+1)} + u^{(t+1)} \implies u^{(t+1)} = 0$. □

Therefore, if the algorithm gets stuck in the inner while loop, $B^{(t+1)} \neq B^{(t)}$. But in each iteration, we add a constraint and there are a finite number of constraints. Therefore, it's not possible to get stuck in the inner while loop.

Lemma 4. *If the algorithm does not terminate in the k^{th} step,*

$$u^{(k)} = 0 \implies (a_{l^{(k)}}^T u^{(k+1)} > 0 \wedge \nabla_f(x^{(k)})^T u^{(k+1)} < 0 \wedge f(x^{(k)} + u^{(k+1)}) < f(x^{(k)}))$$

Proof. Let $g^{(t)} = \nabla_f(x^{(t)})^T = Qx^{(t)} - h$. For any t ,

$$\begin{aligned}
(u^{(t)}, \mu^{(t)}) &= \text{eqsolve}(Q, h - Qx^{(t)}, A_{B^{(t)}}, 0) \\
&\implies \begin{bmatrix} Q & -A_{B^{(t)}}^T \\ A_{B^{(t)}} & 0 \end{bmatrix} \begin{bmatrix} u^{(t)} \\ \mu^{(t)} \end{bmatrix} = \begin{bmatrix} h - Qx^{(t)} \\ 0 \end{bmatrix} \\
&\implies Qu^{(t)} - A_{B^{(t)}}^T \mu^{(t)} = h - Qx^{(t)} = -g^{(t)} \wedge A_{B^{(t)}} u = 0 \\
&\implies \left(g^{(t)} + Qu^{(t)} = A_{B^{(t)}}^T \mu^{(t)} = \sum_{i \in B^{(t)}} \mu_i^{(t)} a_i \right) \wedge (\forall i \in B^{(t)}, a_i^T u = 0)
\end{aligned}$$

For notational convenience, let $l = l^{(k)}$. Since $u^{(k)} = 0, x^{(k+1)} = x^{(k)} \implies g^{(k+1)} = g^{(k)}$ and $B^{(k+1)} = B^{(k)} - \{l\}$. Since the algorithm didn't terminate, $\mu_l^{(k)} < 0$.

$$g^{(k)} = g^{(k)} + Qu^{(k)} = \sum_{i \in B^{(k)}} \mu_i^{(k)} a_i = \mu_l^{(k)} a_l + \sum_{i \in B^{(k+1)}} \mu_i^{(k)} a_i$$

$$g^{(k+1)} + Qu^{(k+1)} = \sum_{i \in B^{(k+1)}} \mu_i^{(k+1)} a_i$$

On subtracting these 2 equations, we get

$$Qu^{(k+1)} = -\mu_l^{(k)} a_l + \sum_{i \in B^{(k+1)}} (\mu_i^{(k+1)} - \mu_i^{(k)}) a_i$$

$\forall i \in B^{(k+1)}, a_i^T \mu_i^{(k+1)} = 0$. Therefore,

$$\begin{aligned} & (u^{(k+1)})^T Qu^{(k+1)} \\ &= -\mu_l^{(k)} a_l^T u^{(k+1)} + \sum_{i \in B^{(k+1)}} (\mu_i^{(k+1)} - \mu_i^{(k)}) a_i^T u^{(k+1)} \\ &= -\mu_l^{(k)} a_l^T u^{(k+1)} \\ &\implies a_l^T u^{(k+1)} = \frac{(u^{(k+1)})^T Qu^{(k+1)}}{-\mu_l^{(k)}} > 0 \quad (Q \text{ is PD and } \mu_l^{(k)} < 0) \end{aligned}$$

$$\begin{aligned} g^{(k)} + Qu^{(k+1)} &= \sum_{i \in B^{(k+1)}} \mu_i^{(k+1)} a_i \\ \implies g^{(k)T} u^{(k+1)} + u^{(k+1)T} Qu^{(k+1)} &= \sum_{i \in B^{(k+1)}} \mu_i^{(k+1)} a_i^T u^{(k+1)} = 0 \\ \implies g^{(k)T} u^{(k+1)} &= -u^{(k+1)T} Qu^{(k+1)} \end{aligned}$$

$$\begin{aligned} & f(x^{(k)} + u^{(k+1)}) \\ &= f(x^{(k)}) + g^{(k)T} u^{(k+1)} + \frac{1}{2} u^{(k+1)T} Qu^{(k+1)} \quad (\text{Taylor series}) \\ &= f(x^{(k)}) - \frac{1}{2} u^{(k+1)T} Qu^{(k+1)} \\ &< f(x^{(k)}) \quad (Q \text{ is PD}) \end{aligned}$$

□

Corollary 4.1. $u^{(k)} = 0 \implies u^{(k+1)} \neq 0$

Lemma 5. $\alpha^{(t)} > 0 \implies f(x^{(t+1)}) < f(x^{(t)})$

Proof. For $\alpha^{(t)}$ to exist, $u^{(t)}$ must be non-zero. Therefore, $x^{(t)} \neq x^{(t)} + u^{(t)}$.

$$u^{(t)} = \underset{A_{B^{(t)}} u=0}{\operatorname{argmin}_u} f(x^{(t)} + u) \implies x^{(t)} + u^{(t)} = \underset{(A_{B^{(t)}})_x = (b_{B^{(t)}})}{\operatorname{argmin}_x} f(x)$$

Since $x^{(t)}$ satisfies $A_{B^{(t)}}x = b_{B^{(t)}}$, it is a feasible solution to the above optimization problem. However, $x^{(t)} + u^{(t)}$ is the optimal solution and $x^{(t)} \neq x^{(t)} + u^{(t)}$. Since f is a strictly convex function, it has a unique global minimum. Therefore, $f(x^{(t)} + u^{(t)}) < f(x^{(t)})$.

$$\begin{aligned}
f(x^{(t+1)}) &= f(x^{(t)} + \alpha^{(t)}u^{(t)}) \\
&= f((1 - \alpha^{(t)})x^{(t)} + \alpha^{(t)}(x^{(t)} + u^{(t)})) \\
&\leq (1 - \alpha^{(t)})f(x^{(t)}) + \alpha^{(t)}f(x^{(t)} + u^{(t)}) \\
&< (1 - \alpha^{(t)})f(x^{(t)}) + \alpha^{(t)}f(x^{(t)}) \quad (f(x^{(t)} + u^{(t)}) < f(x^{(t)}) \text{ and } \alpha^{(t)} > 0) \\
&= f(x^{(t)})
\end{aligned}$$

□

Lemma 6. Let $N^{(t)} = \{i : a_i^T u^{(t)} < 0\}$. Then

$$\begin{aligned}
B^{(t)} \cap N^{(t)} &= \{\} & \mathcal{A}(x^{(t)}) \cap N^{(t)} &= (\mathcal{A}(x^{(t)}) - B^{(t)}) \cap N^{(t)} \\
\alpha^{(t)} = 0 &\iff \mathcal{A}(x^{(t)}) \cap N^{(t)} \neq \{\} &\iff I^{(t)} = \mathcal{A}(x^{(t)}) \cap N^{(t)} \wedge I^{(t)} &\neq \{\}
\end{aligned}$$