

Machine Learning

Instructor: Prof. Ganesh Ramakrishnan

Projection methods

- So far, we have dealt with simple projections during SMO and the general decomposition method
 - ▶ We considered $\alpha_i y_i + \alpha_j y_j = \text{constant}$, and solved a quadratic optimization problem for α_i and α_j
 - ▶ We then projected $(\alpha_i, \alpha_j) \rightarrow [0, C]^2$
- We will now 'scale up' these projections
- In active set methods, the working set changes slowly. Projection methods can solve bound constrained optimization problems with large changes in the working set at each iteration.

The general SLBQP problem

- We define the general SLBQP (**S**ingle **L**inear equality constrained **B**ounded **Q**uadratic **P**rogram) problem as

$$\min f(x) = \frac{1}{2}x^T Ax - c^T x$$

s.t.

- ▶ $l_i \leq x_i \leq u_i, \forall i$
- ▶ $a^T x = b$

These constraints form the non-empty closed convex set Ω

- This can be considered a generalization of the SVM dual objective where
 - ▶ $x_i \equiv \alpha_i$
 - ▶ $A_{ij} \equiv Q_{ij} = y_i y_j K(x_i, x_j)$
 - ▶ $c_i \equiv 1$
 - ▶ $l_i \equiv 0$
 - ▶ $u_i \equiv C$
 - ▶ $a_i \equiv y_i$
 - ▶ $b \equiv 0$

Overview

- We can find Δx as the change in x along some steepest descent direction of f without constraints
- Thus, let $x_u^{k+1} = x^k + \Delta x$ be the working set that reduces $f(x)$ without constraints (unbounded)
- To find the constrained working set, we project x_u^{k+1} onto Ω to get x^{k+1}

$$\Omega = \bigcap_{i=1}^m \{x \mid g_i(x) \leq 0\}$$

Algo: Initialise: $x_u^{(0)}$, $x_p^{(0)} = P_\Omega(x_u^{(0)})$
 until convergence, $x_u^{(k+1)} = x_u^{(k)} - t^{(k)} \nabla f(x_u^{(k)})$
 $x_p^{(k+1)} = P_\Omega(x_u^{(k+1)})$

- To project x_u onto the non-empty closed convex set Ω to get the projected point x_p , we solve:

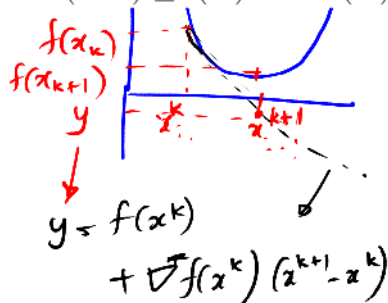
$$x_p = P_\Omega(x_u) = \operatorname{argmin}_{z \in \Omega} \|x_u - z\|_2^2$$

- That is, the projected point x_p is the point in Ω that is the closest to the unbounded optimal point x_u if Ω is a non-empty closed convex set

Recall: If g_i 's are lower semi-cts then Ω is closed convex & a unique x_p is guaranteed to exist

Descent direction for a convex function

- For a descent in a convex function f , we must have $f(x^{k+1}) \geq$ Value at x^{k+1} obtained by linear interpolation from x^k
- i.e. $f(x^{k+1}) \geq f(x^k) + \nabla^T f(x^k)(x^{k+1} - x^k)$



- Thus, for Δx^k to be a descent direction, it is necessary that $\nabla^T f(x^k) \Delta x^k \leq 0$
 (where $\Delta x^k = x^{k+1} - x^k$)

Q: Is $\nabla^T f(x^k) (P_{\Omega}(x^k - t^k \nabla f(x^k)) - x^k) \leq 0$?

We want that the point obtained after the projection of x_u^{k+1} to be a descent from x^k for the function f

$$\nabla f(x^k) \cdot \Delta x_p \leq 0$$

(where $\Delta x_p = P_{\Omega}(x_u^{k+1}) - x^k$)

} This is only necessary ... for complete convergence

Ref Nemirovski Sec 5.3.1 ←

- **Claim:** If $P_{\Omega}(x)$ is a projection of x , then

} Actually claim

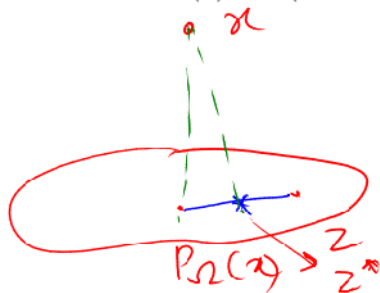
$$(z - P_{\Omega}(x))^{\top} (x - P_{\Omega}(x)) \leq 0, \forall z \in \Omega$$

- That is, the angle between $(z - P_{\Omega}(x))$ and $(x - P_{\Omega}(x))$ is obtuse (or right-angled for the projected point), $\forall z \in \Omega$



Proof for $\langle z - P_{\Omega}(x), x - P_{\Omega}(x) \rangle \leq 0$

- To be more general, let us consider an inner product $\langle a, b \rangle$ instead of $a^T b$
- Let $z^* = (1 - \alpha)P_{\Omega}(x) + \alpha z$, for some $\alpha \in (0, 1)$, and $z \in \Omega$
 $\implies z^* = P_{\Omega}(x) + \alpha(z - P_{\Omega}(x)), z^* \in \Omega$



- Since $P_{\Omega}(x) = \operatorname{argmin}_{z \in \Omega} \|x - z\|_2$,
 $\|x - P_{\Omega}(x)\|^2 \leq \|x - z^*\|^2$

$$\begin{aligned}
& \|x - z^*\|^2 \\
&= \left\| x - (P_\Omega(x) + \alpha(z - P_\Omega(x))) \right\|^2 \\
&= \|x - P_\Omega(x)\|^2 + \alpha^2 \|z - P_\Omega(x)\|^2 - \underbrace{2\alpha \langle x - P_\Omega(x), z - P_\Omega(x) \rangle}_{\geq 0} \\
&\geq \|x - P_\Omega(x)\|^2 \\
&\implies \underbrace{\langle x - P_\Omega(x), z - P_\Omega(x) \rangle}_{\text{LHS}} \leq \frac{\alpha}{2} \|z - P_\Omega(x)\|^2, \forall \alpha \in (0, 1)
\end{aligned}$$

- Thus, the LHS can either be 0 or a negative value. Any positive value of the LHS will lead to a contradiction for some small $\alpha \rightarrow 0$
- Hence, we proved that $\langle z - P_\Omega(x), x - P_\Omega(x) \rangle \leq 0$

- We can also prove that if $\langle x - x^*, z - x^* \rangle \leq 0, \forall z \in \Omega$ s.t. $z \neq x^*$, and $x^* \in \Omega$, then

$$x^* = P_{\Omega}(x) = \operatorname{argmin}_{\bar{z} \in \Omega} \|x - \bar{z}\|_2^2$$

- Consider $\|x - z\|^2 - \|x - x^*\|^2$

$$= \|x - x^* + (x^* - z)\|^2 - \|x - x^*\|^2$$

$$= \|x - x^*\|^2 + \|z - x^*\|^2 - 2 \langle x - x^*, z - x^* \rangle - \|x - x^*\|^2$$

$$= \|z - x^*\|^2 - 2 \langle x - x^*, z - x^* \rangle$$

$$> 0$$
- $\implies \|x - z\|^2 > \|x - x^*\|^2, \forall z \in \Omega$ s.t. $z \neq x^*$
- This proves that $x^* = P_{\Omega}(x)$

References

- Yu-Hong Dai, Roger Fletcher. New algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds. <http://link.springer.com/content/pdf/10.1007%2Fs10107-005-0595-2.pdf>