# MATH 565: Lecture 19 (03/24/2026)

Today:
* inequality constraints
* box constraints
* Lagrangian relaxation

## Inequality Constraints

Recall We are studying $\min \{ F(\bar{\omega}) \mid \bar{\omega} \in C \}$. When $C = \{ \bar{\omega} \mid A\bar{\omega} = \bar{b} \}$, we could "eliminate variables". But we cannot directly do that when $C = \{ \bar{\omega} \mid A\bar{\omega} \le \bar{b} \}$.

Q. How do we handle inequality constraints in $C$?

### A. Conditional Gradient Optimization using LP

Set

$$(LP) \left\{ \bar{\omega}_{t+1} = \underset{\bar{\omega}}{\operatorname{argmin}} \ \overbrace{F(\bar{\omega}_t) + [\nabla F(\bar{\omega}_t)]^T (\bar{\omega} - \bar{\omega}_t)} \quad \text{s.t. } A\bar{\omega} \le \bar{b} \right.$$

first Taylor expansion of $F(\bar{\omega})$ at $\bar{\omega}_t$; linear function

This is a linear program (LP), and can be solved efficiently.

But $\bar{\omega}_{t+1}$ may not minimize $F(\bar{\omega})$. So, we set

$$\bar{q}_t = \bar{\omega}_{t+1} - \bar{\omega}_t \quad \text{(descent direction)}$$

and then update

$$\bar{\omega}_{t+1} \leftarrow \bar{\omega}_t + \alpha_t \bar{q}_t$$

where $\alpha_t$ is a line search parameter chosen to ensure feasibility and optimality of $F(\bar{\omega})$.

See LO4ML for sequential LP and sequential QP approaches.

## B. Special Case: Box Constraints

In many ML problem settings, the inequality constraints are simple box constraints of the form

$$C = \{\bar{w} \mid l_i \leq w_i \leq u_i,\ i=1,\dots,d\}.$$

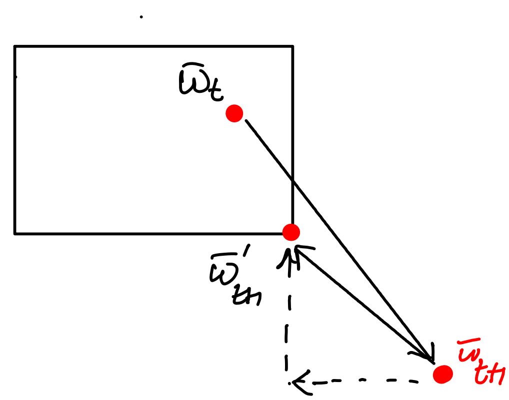
IDEA: Take a gradient descent step and project back to the boundary of the box if landed outside. This projection can be implemented one dimension at a time.

Here is the general algorithm.

1. $\bar{w}^{t+1} \leftarrow \bar{w}^{t} - \alpha_t \nabla F(\bar{w}^{t})$ (gradient descent step)

2. If $w_i < l_i$, $w_i \leftarrow l_i$. else if $w_i > u_i$, $w_i \leftarrow u_i$. (project back to boundary of box if landed outside — one dimension at a time).

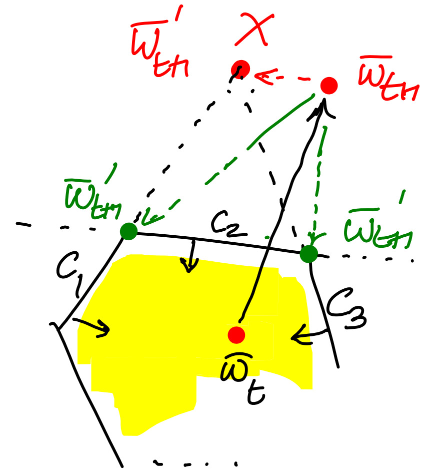
Projecting $\bar{w}_{t+1}$, which is outside the box, back to the boundary of the box so that violations of the box in each dimension are fixed.




Note that the constraints defining $C$ here are naturally linearly independent (LI) — we'll try to use this fact when generalizing to $C$ that are polytopes (and not boxes).

## C. General $A\bar{w} \le \bar{b}$

Let's consider how we may generalize the projection approach from the box case to $C = \{\bar{w} \mid A\bar{w} \le \bar{b}\}$. In the box case, we have $d$ LI constraints in $\mathbb{R}^d$, and hence projecting back to Box is always unambiguous. But we usually have (many) more than $d$ constraints when considering $A\bar{w} \le \bar{b}$: see 2D instance as shown here.




Consider the constraints $c_1, c_2, c_3$: any pair of them are LI, but all three of them together are not LI in 2D. If we pick $\{c_1, c_2\}$ or $\{c_2, c_3\}$ as a subset of LI constraints that are violated, projecting back to either point of intersection will work. But if we choose $\{c_1, c_3\}$, their point of intersection is not feasible.

In the general case, we could have more than $d$ constraints that are violated, and some choices of subsets of $d$ LI constraints (among those violated) for projecting back may not give feasible points (i.e., $\bar{w}_{t+1} \notin C$ is possible).

But if all rows of $A$ in $C = \{\bar{w} \mid A\bar{w} \le \bar{b}\}$ are LI to start with, then we can project in all cases, since any subset of constraints that are violated will also be LI.

Here is the general approach in this case:

1. $\bar{\omega}_{t+1} = \bar{\omega}_t - \alpha_t \nabla F(\bar{\omega}_t)$ gradient descent step

2. if $\bar{\omega}_{t+1} \notin C = \{\bar{\omega} \,|\, A\bar{\omega} \le \bar{b}\}$,
let $A_v \bar{\omega} \le \bar{b}_v$ be the subset of constraints violated by $\bar{\omega}_{t+1}$ rows of $A_v$ are LI here

3. Update $\bar{\omega}_{t+1}$ as
$$\bar{\omega}'_{t+1} \leftarrow \bar{\omega}_{t+1} + \underbrace{A_v^T (A_v A_v^T)^{-1}} [\bar{b}_v - A_v \bar{\omega}_{t+1}]$$
right inverse formula (see Lecture 18), but now applied to the slack $\bar{b}_v - A_v\bar{\omega}_{t+1}$ (instead of $\bar{b}_v$, which would've been used if it were $A_v\bar{\omega} = \bar{b}_v$).

Check: $A_v \bar{\omega}'_{t+1} = \cancel{A_v \bar{\omega}_{t+1}} + \underbrace{A_v A_v^T (A_v A_v^T)^{-1}}_{I} [\bar{b}_v - \cancel{A_v \bar{\omega}_{t+1}}]$
$= \bar{b}_v$ !

4. $t \to t+1$ if $\|\bar{\omega}'_{t+1} - \bar{\omega}_t\| > \epsilon$ (not converged yet)

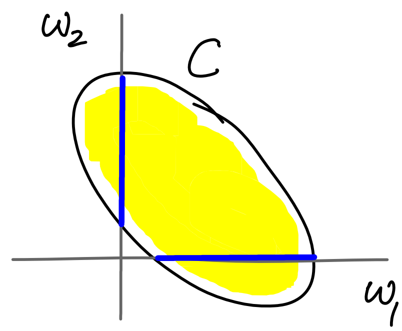
But requiring all rows of $A$ to be LI may be too strict. In practice, rows of $A_v$ are often LI (even when those of $A$ are not). We could also choose $\alpha_t$ (line search parameter) to ensure that the rows of $A_v$ are LI (in each iteration).

# D. Primal Coordinate Descent (CD) for $A\bar{w} \le \bar{b}$ constraints

Recall coordinate descent (CD) for unconstrained loss minimization from Lecture 8...

In $\min\{F(\bar{w}) \mid \bar{w} \in C\}$, when $F(\bar{w})$ is a convex function and $C$ is a convex set, CD can be quite effective.

This is because the intersection of a convex $C$ with any coordinate axis will be a single line segment (assuming $C$ is closed and bounded).




Hence, when all but one variable $w_j$ is fixed, the minimization problem in the $j^{th}$ coordinate can be handled similar to how we handle box constraints.

See LO4ML for details...

# Lagrangian Relaxation and Duality

This is a versatile and quite general method to handle constraints, and uses the fundamental concept of duality.

Consider $\pi^* = \min F(\bar{\omega})$ (P) primal problem

s.t. $f_i(\bar{\omega}) \leq 0, \quad i=1,\dots,m$

This is the primal problem, denoted as (P). We let $\pi^*$ be the optimal objective function value for an optimal solution $\bar{\omega}^*$, while $\pi = F(\bar{\omega})$ for any feasible $\bar{\omega}$.

Then the **Lagrangian relaxation** of (P) using the non-negative multipliers $\bar{\alpha} = [\alpha_1 \cdots \alpha_m]^T$ is

$$L(\bar{\alpha}) = \min_{\bar{\omega}} F(\bar{\omega}) + \sum_{i=1}^{m} \alpha_i f_i(\bar{\omega})$$

The minimization here is done only over $\bar{\omega}$; note that $\bar{\alpha} \geq \bar{0}$ is a vector parameter (assumed to be fixed).

This is a relaxation of (P) since the constraints are not explicitly imposed any more. But, if constraint $i$ is violated, $f_i(\bar{\omega}) > 0$, and hence the term $\alpha_i f_i(\bar{\omega})$ increases $L(\bar{\alpha})$, which works against the minimization objective.

But we can show that $L(\bar{\alpha})$ provides a lower bound for $\pi^*$, the optimal objective function value of (P).

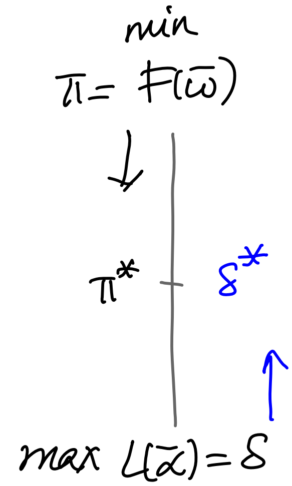
Let $\bar{\omega}^*$ be optimal for (P), giving $\pi^* = F(\bar{\omega}^*)$.

$$\Rightarrow \quad f_i(\bar{\omega}^*) \leq 0 \quad \forall i \quad (\text{as } \bar{\omega}^* \text{ is feasible})$$

$$\Rightarrow \quad \alpha_i f_i(\bar{\omega}^*) \leq 0 \quad \forall i, \quad \text{as} \quad \alpha_i \geq 0.$$

$$\Rightarrow \quad L(\bar{\alpha}) \leq F(\bar{\omega}^*) + \underbrace{\sum_{i=1}^{m} \alpha_i f_i(\bar{\omega}^*)}_{\leq 0}$$

$$\leq F(\bar{\omega}^*)$$




$\Rightarrow$ $L(\bar{\alpha})$ gives a lower bound for $\pi^*$ $\forall \bar{\alpha} \geq \bar{0}$. With $\delta = L(\bar{\alpha})$ for any $\bar{\alpha} \geq \bar{0}$. We can try to make this lower bound as tight as possible by maximizing $L(\bar{\alpha})$ over $\bar{\alpha} \geq \bar{0}$. This gives the dual problem:

$$\text{(D)} \quad \delta^* = \max_{\bar{\alpha} \geq \bar{0}} L(\bar{\alpha})$$

dual problem

Let $\bar{\alpha}^*$ be optimal for (D), and let $\delta^* = L(\bar{\alpha}^*)$ be the maximum objective function value. Then we get that

$$\delta^* = L(\bar{\alpha}^*) \leq \pi^* = F(\bar{\omega}^*)$$

This result is called **weak duality** (weak since it is $\leq$ and not $=$).

we get strong duality if we have $\delta^* = \pi^*$.

Combining the two problem formulations, we can write (D) as

$$\delta^* = \max_{\bar{\alpha} \geq \bar{0}} \min_{\bar{w}} \left\{ F(\bar{w}) + \sum_{i=1}^{m} \alpha_i f_i(\bar{w}) \right\}.$$

This is a max-min optimization problem, and that order is important — you cannot flip the order of max and min in general...