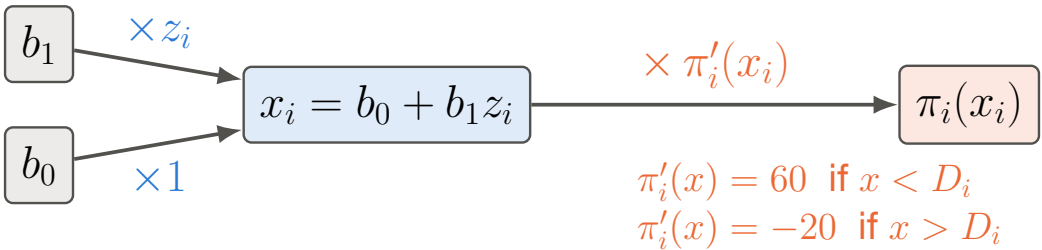
Policies that satisfy constraints

208424 Lecture 16

Lecture 15: learning the policy directly

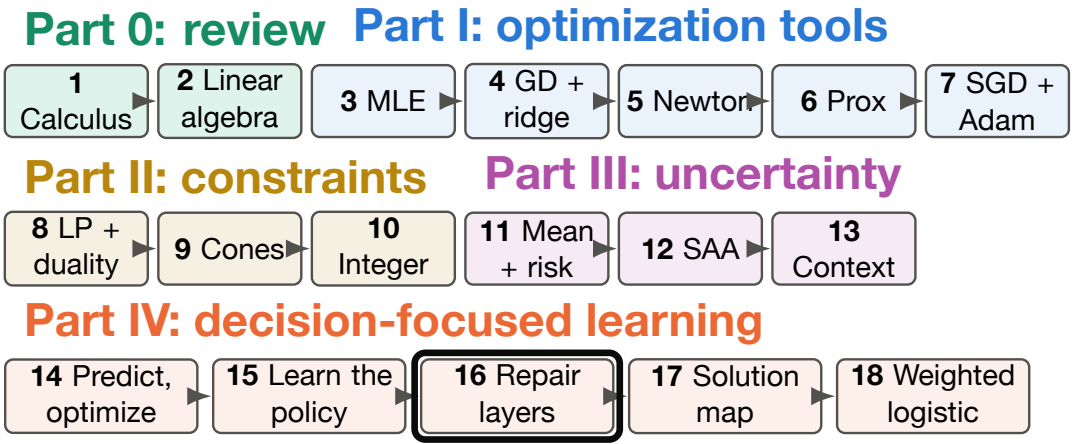


In Lecture 15, we maximized the average profit $\hat{V}$.

SGD used the chain rule for $\partial\pi_i/\partial b_0$ and $\partial\pi_i/\partial b_1$.

Today, two cafés share one truck.

Course map


Part 0: review
Part I: optimization tools
1 Calculus
2 Linear algebra
3 MLE
4 GD + ridge
5 Newton
6 Prox
7 SGD + Adam
Part II: constraints
Part III: uncertainty
8 LP + duality
9 Cones
10 Integer
11 Mean + risk
12 SAA
13 Context
Part IV: decision-focused learning
14 Predict, optimize
15 Learn the policy
16 Repair layers
17 Solution map
18 Weighted logistic

Lecture 16:
Policies that satisfy constraints

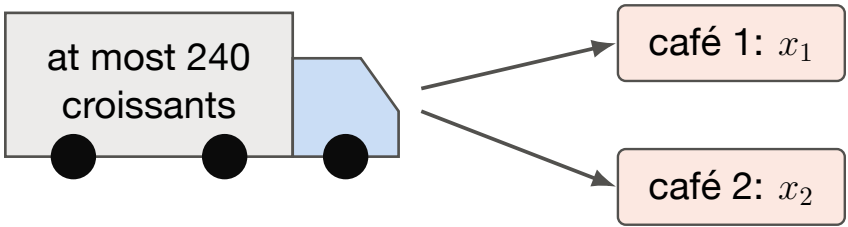


Photo: Mj-bird, CC BY-SA 3.0, Wikimedia Commons

Two cafés get croissants from one truck.

On busy days, the truck cannot carry every order.

Two cafés and one truck

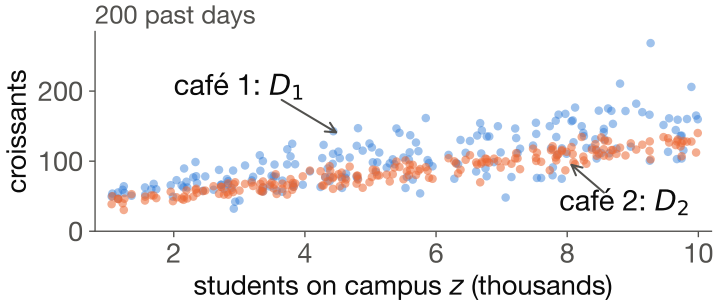


One truck brings croissants to two cafés.

Today, the truck holds at most 240 croissants.

So the orders must satisfy $x_1 + x_2 \leq 240$.

Demand at the two cafés

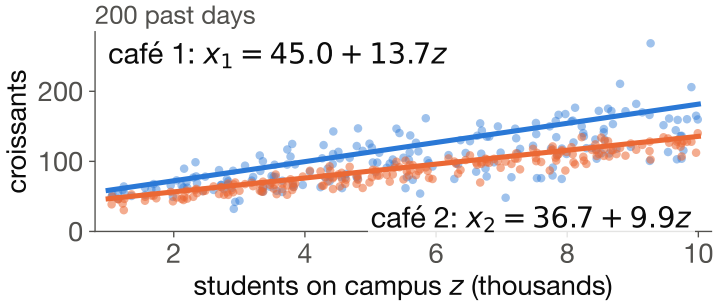


On each past day, we know z, D_1 and D_2.

At both cafés, demand increases with z.

Café 2's demand has a smaller SD than café 1's.

A policy for each café

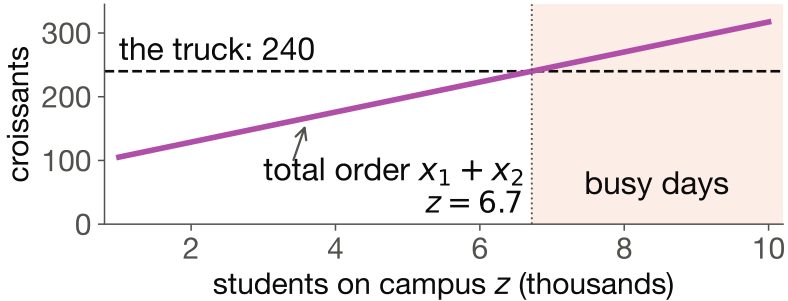


Each café maximizes its own profit on past days.

So each café learns its own linear policy.

Both policies ignore the truck.

The total order

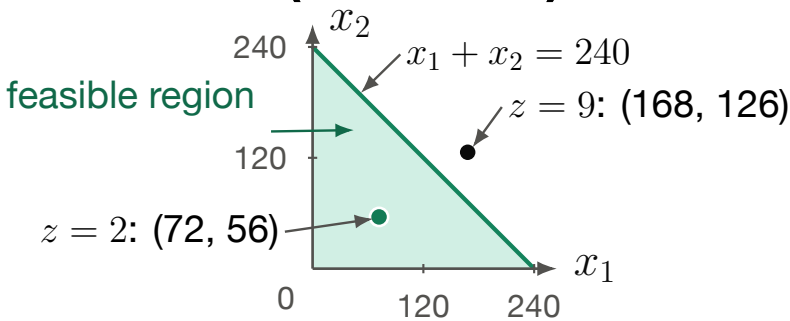


The total order $x_1 + x_2$ increases with z.

When z is above 6.7, the total is more than 240.

On these days, the orders do not fit in the truck.

Feasible orders (Lecture 8)

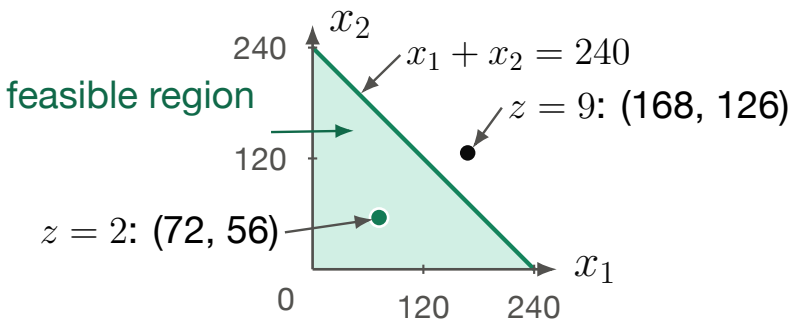


Feasible orders satisfy every constraint.

The feasible region is a triangle.

On a busy day, the orders are not feasible.

Summary: orders and the truck



Each café's policy ignores the truck.

On busy days, the orders are not feasible.

Checkpoint 1

café 1: $x_1 = 30 + 15z$

café 2: $x_2 = 25 + 10z$

the truck holds at most 240

For which z are the orders feasible?

Checkpoint 1

café 1: $x_1 = 30 + 15z$

café 2: $x_2 = 25 + 10z$

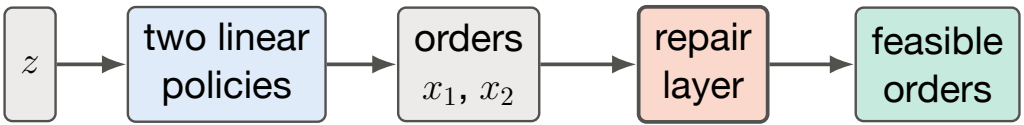
the truck holds at most 240

For which z are the orders feasible?

The total order $55 + 25z$ must be at most 240.

So the orders are feasible when $z \leq 7.4$.

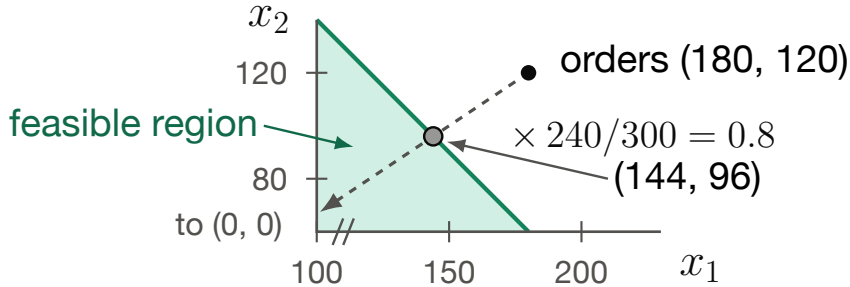
A repair layer



A repair layer makes the orders feasible.

A repair layer does not change feasible orders.

Scaling

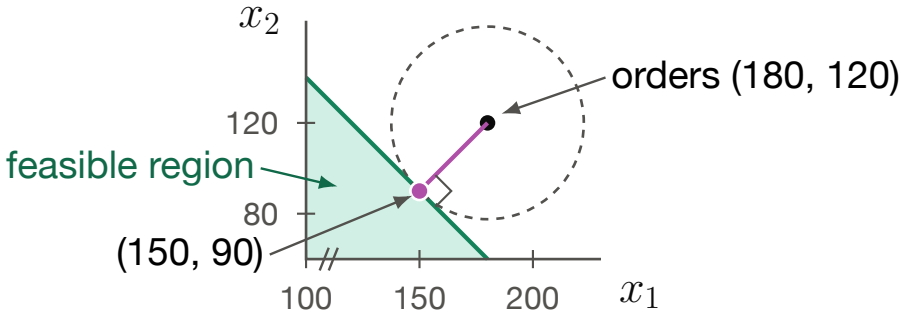


Scaling gives $y_k = 240x_k/(x_1 + x_2)$ for $k = 1, 2$.

We scale only when $x_1 + x_2 > 240$.

So $(180, 120)$ becomes $(y_1, y_2) = (144, 96)$.

Projection



The projection is the nearest feasible point.

The projection of (180, 120) is (150, 90).

We take $(180 + 120 - 240)/2 = 30$ from each order.

Scaling and projection

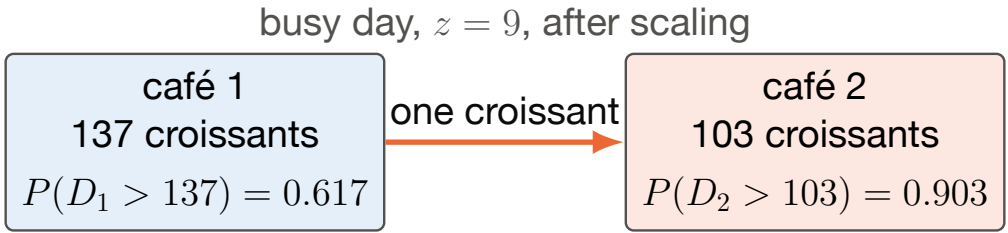
	café 1	café 2
orders	180	120
after scaling ($\times 0.8$)	144 (-36)	96 (-24)
after projection	150 (-30)	90 (-30)

Scaling makes each order 20% smaller.

Projection takes 30 croissants from each order.

Both repair layers depend only on the orders.

Moving one croissant

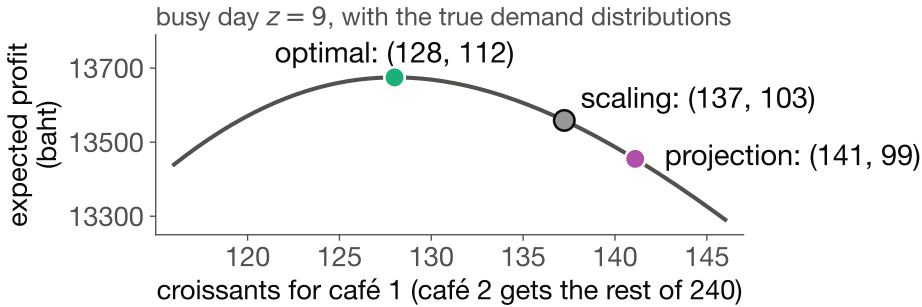


We move one croissant from café 1 to café 2.

The croissant is more likely to sell at café 2.

So the expected profit increases by about 23 baht.

The optimal allocation

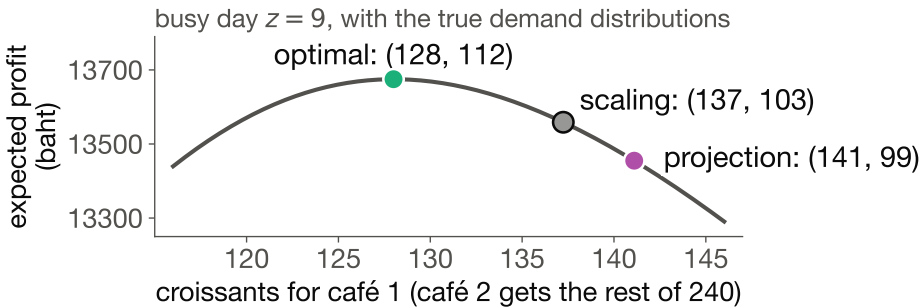


An allocation divides the truck between the cafés.

The optimal allocation is (128, 112).

At this allocation, $P(D_1 > 128) = P(D_2 > 112)$.

Summary: repair layers



Scaling and projection make the orders feasible.

The optimal allocation depends on the demand.

Checkpoint 2

orders: $x_1 = 200,\ x_2 = 120$

the truck holds at most 240

Find the orders after scaling.

Find the orders after projection.

Checkpoint 2

orders: $x_1 = 200,\ x_2 = 120$

the truck holds at most 240

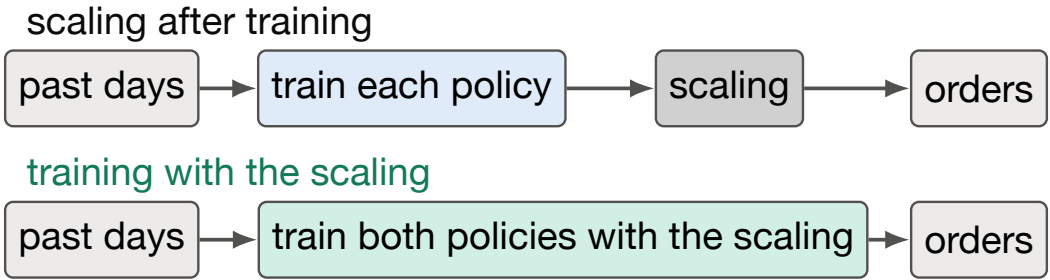
Find the orders after scaling.

Find the orders after projection.

With $240/320 = 0.75$, scaling gives (150, 90).

With $(320 - 240)/2 = 40$, projection gives (160, 80).

When to add the repair layer



We can add the scaling after training.

We can also train the policies with the scaling.

Then the policies can learn the allocation.

The average profit with scaling

$$\hat{V} = \frac{1}{n} \sum_{i=1}^{n} \pi_i$$

$\pi_i = 80 \min(D_1, y_1) - 20y_1 + 80 \min(D_2, y_2) - 20y_2$
with the demands and scaled orders of day i

We simulate the past days with the scaled orders.

Now $\hat{V}$ is the average total profit of the two cafés.

A larger x_1 on a busy day

orders	after scaling (the truck is full) y_1 (café 1)	y_2 (café 2)
$x_1 = 180,\ x_2 = 120$	144	96
$x_1 = 181,\ x_2 = 120$	144.32	95.68

When x_1 increases from 180 to 181,

café 1 gets 0.32 more croissants.

So $\partial y_1/\partial x_1 = 0.32$ and $\partial y_2/\partial x_1 = -0.32$.

Partial derivative $\partial \pi_i / \partial x_1$ on a busy day

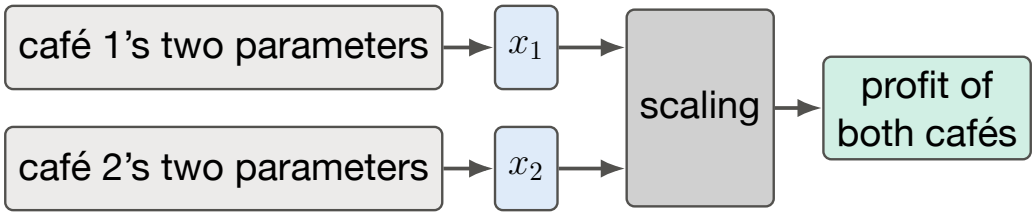
	croissants after scaling	demand	one more croissant (baht)	$\partial y_k / \partial x_1$
café 1	144	160	$+60$	$+0.32$
café 2	96	80	-20	-0.32

Café 1 sold out, and café 2 had croissants left.

Then $\partial \pi_i / \partial x_1 = 60 \times 0.32 + (-20) \times (-0.32) = 25.6$.

So a gradient step increases x_1.

Gradient steps with the scaling

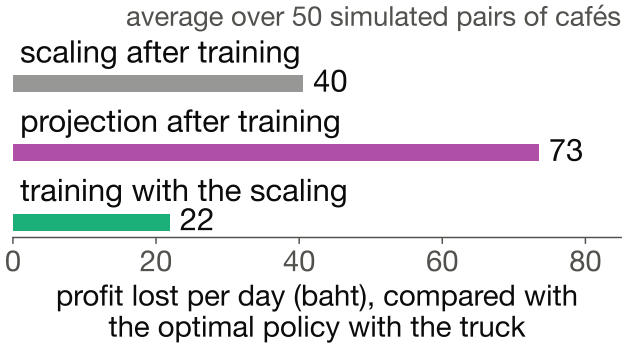


If the truck is not full, the derivatives are Lecture 15's.

The chain rule gives the four partial derivatives.

We take gradient steps to maximize $\hat{V}$.

Profit lost on new days

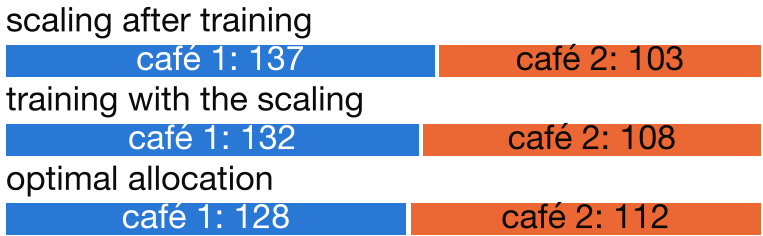


Training with the scaling loses 22 baht per day.

Scaling after training loses 40 baht per day.

The allocation on a busy day

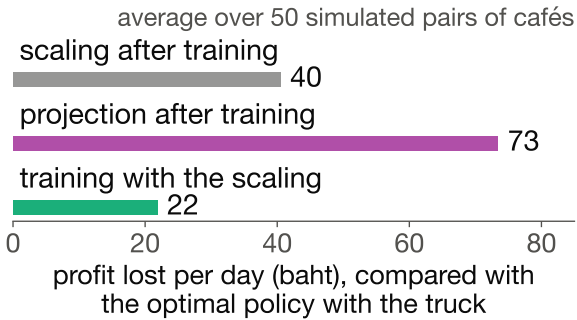
240 croissants on the busy day $z = 9$



On this busy day, training with the scaling gives café 2 more croissants.

So the allocation is closer to the optimal allocation.

Summary: training with a repair layer



Training with the scaling loses the least profit.

We find the partial derivatives with the chain rule.

Checkpoint 3

orders (180, 120)
after scaling (144, 96)

demand on this past day
$D_1 = 160, \;\; D_2 = 110$

Both cafés sold out on this past day.

What is $\partial \pi_i / \partial x_1$?

Checkpoint 3

orders (180, 120)
after scaling (144, 96)

demand on this past day
$D_1 = 160, \;\; D_2 = 110$

Both cafés sold out on this past day.

What is $\partial \pi_i / \partial x_1$?

One more croissant adds 60 baht at each café.

So $\partial \pi_i / \partial x_1 = 60 \times 0.32 + 60 \times (-0.32) = 0.$

Power grids



Photo (a power plant in Iceland): Gretar Ívarsson, edited by Fir0002, public domain, Wikimedia Commons

Chen et al. (2024)

Researchers train a model for power grids.

The model has a scaling repair layer inside.

The model replaces a solver for a known problem.

A warehouse and its stores

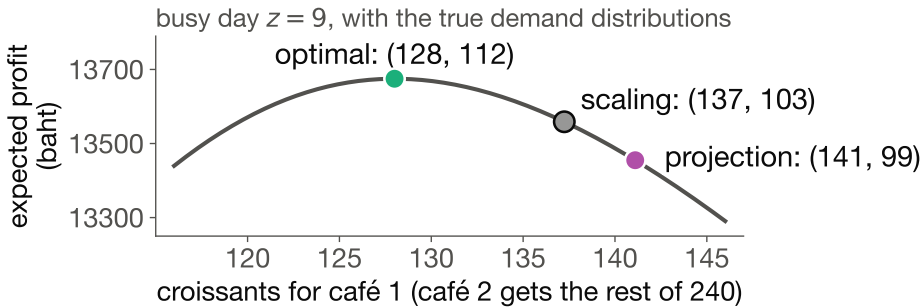


Photo: Mark LS, CC BY-SA 4.0, Wikimedia Commons (cropped)

Alvo et al. (2025)

If the stores order more than the warehouse has,
all orders are scaled down.
The policies are trained with the scaling inside.

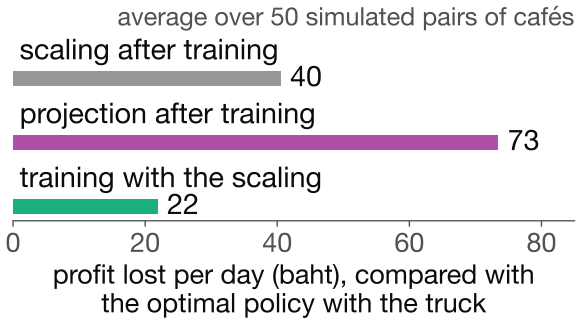
Lecture summary



Scaling and projection make the orders feasible.

The optimal allocation depends on the demand.

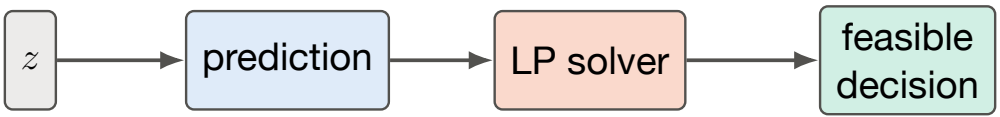
Lecture summary (continued)



Training with the scaling loses the least profit.

We find the partial derivatives with the chain rule.

Next lecture



In Lecture 17, a solver chooses the decision.

We predict the LP's coefficients, then solve the LP.

The solver's decision satisfies every constraint.

Exercises

1 orders (240, 120)
the truck holds at most 240

2 after scaling, café 1 has 150 and café 2 has 90
$P(D_1 > 150) = 0.5$ and $P(D_2 > 90) = 0.8$

3 orders (180, 120), after scaling (144, 96)
demand on this past day: $D_1 = 100$, $D_2 = 150$

1. Find the orders after scaling and projection.
2. Should a croissant move from café 1 to café 2?
3. Find $\partial \pi_i / \partial x_1$ on this past day.