Problem 5:

Stochastic Gradient Decent (SGD) method is suggested, which can be particularly useful in large scale problems.

SGD tries to lower the computation required per iteration (although the number of iterations may increase for convergence of the problem). In SGD we can select a random time step k where we can carry out the learning algorithm to correct the dual price vector y. The optimization problem is,

A math symbols on a white background

Description automatically generated

The derivative of the function is

A black text on a white background

Description automatically generated



Algorithm:

1. For k=1, initialize the dual price vector y as 0. (as mentioned in the Problem)
2. For k=2, use β = 1/√(2). m is the time step chosen for calculating dual price.

**y***k+1* = **y**m - β-1*∇ f* (**y**m).

**y***k+1* = **0** – 1/√2(*∇ f* (**0**)). This will provide the next set of dual prices. Decision for next step can be made with the help of this dual price. Use this to calculate the revenue. And compare this information with ground truth price.

1. For k=3, use β = 1/√(3). But for calculating the dual prices pick out any time step information **m** between k=2 and k=3 randomly and use that to calculate the next set of dual prices with the formula mentioned in step 2.
2. For all following time steps k chose random time step **m** between 1 and k and chose β = 1/√(k). Use the formula,

**y***k+1* = **y**m - β-1*∇ f* (**y**m).

(for eg if k = 50, chose any time step between 2-50 and calculate the dual prices with β = 1/√(50).)

* Rather than updating the dual price at every time step we can also update in every 10/20/30 time steps. (using the mini batch concept will reduce the variance of the stochastic gradient).

1. The true minimizer is approached when f(**x***k*) - f(**x**\*) ≤ 2β||**x**0 - **x**\*||/k.

Observations:

The Step size that we pick up in SGD decides the problems convergence to the optimum solution. The smaller the step size the problem appears less stochastic, and it will lead to huge number of iterations with very less speed of convergence. Picking up the step size is crucial for the problem. As stochastic model gives a quick convergence in the beginning hence it is particularly useful for big data problems. Hence this works great on unseen data.

Convergence to ground truth price

As the Stochastic approach to gradient method although helps in rapid convergence in the beginning but it fluctuates a lot near the optimal solution. But the amount of noise can be controlled by the variance of the stochastic gradient. The smaller the variance the more stochastic gradient is close to the true gradient.