Differential Privacy Assignment 1

Mart Veldkamp

I. Introduction

In this assignment we were tasked with implementing the Distributed Projected Gradient Method and Private Distributed Projected Gradient Method for a group of n people.

Firs,t simulation without any privacy protection will be made; this will be a baseline for non-private systems.

After this there will be extensive simulation done for in- and decreasing privacy.

And a conclusion will be given at the end of every simulation explaining the results.

II. Variable list

- $T = 50$, this is the total number of iterations.
- $q = 0.6$, with $q \in (0, 1)$.
- $c = 1$, with $c > 0$.
- $p = 0.9$, with $p \in (q, 1)$.
- $v = [0.1, 0.5, 0.4, 0.2]$, these are all the private values.
- $x \in \mathbb{R},\ -1 \leq x \leq 1$, randomly generated.
- $\chi =$ list of x with size n big.
- $n = 4$, this is the total number of elements within v.
- ϵ = 0.001, your privacy value (lower = more privacy).
- $\gamma = cq^t$, the step size.
- C_2 = 3, we got this after solving $C_2 \geq ||\nabla f_i(x)||$
- $f_i(x) = ||x - v_i||^2$
- $b_t = 2C_2\sqrt{n}\ \frac{cp}{\epsilon(p-q)}p^t$, which is the noise, this is needed for the Laplace mechanism.
- $A = \frac{1}{n}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$, The connectivity matrix.

We will call the values within the list x users, and the matrix A is chosen based on VIII, which a fully connected matrix suffices.

When talking about reaching consensus; this formally means $\sum_{i=1}^{n} v_i/n$. And in all our cases is equal to 0.3. Informally; if users reach consensus, they find the average of the list χ.

Note: These variables were chosen for the first implementation, if stated otherwise, the same variables were chosen.

III. Implementation Algorithm 1

As mentioned before, for the first tests, we used the base variable list. This algorithm did not add the Laplace noise. So we expect it to converge to the consensus (0.3).

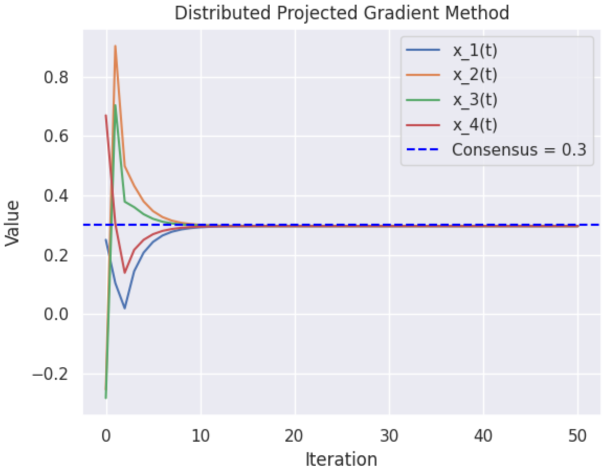

Fig. 1. No private Distributed projected gradient. This is our baseline for non-private communication.

IV. Implementation Algorithm 2

For our second implementation, we are adding the Laplace mechanism. Keeping all the variables in the list the same, note that we only now use the variables ϵ and b_t. Which are used in noise calculations, as can be seen.

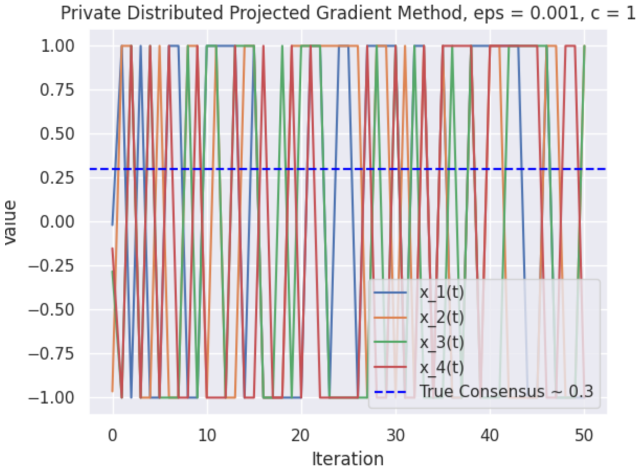

Fig. 2. Low ϵ private gradient. Which gives a lot of privacy, but little insight.

As can be seen from 2, a low ϵ made it so that the consensus never reached the expected value. This isn't bad, it just means that it's private. Which tells you little about the actual data.

We can change this by setting ϵ to 10^5. What we expect is less private. But also gives us our expected "True Consensus", as can be seen from 3. Which reduces privacy, but gives our expected consensus.

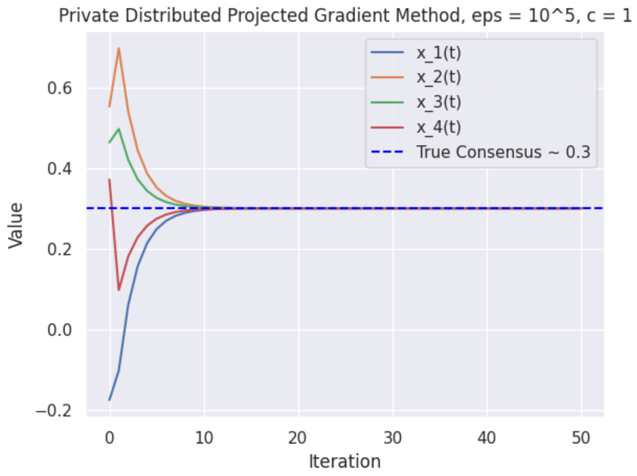

Fig. 3. Test with a high value of ϵ, which by definition means less private.

Another variable we can tune is c, which is both correlated to changing the step size γ and our noise b_t. Decreasing c from $1 \rightarrow 0.5$ or lower will therefore decrease the step size, and decrease privacy (minimally, due to p^t being dominant as $t \rightarrow \infty$).

Increasing c on the other hand will therefore increase step size, and increase privacy (still minimal). In 4 an ϵ and c were given where the users were given some information, but still were moderately private for $t \rightarrow 50$.

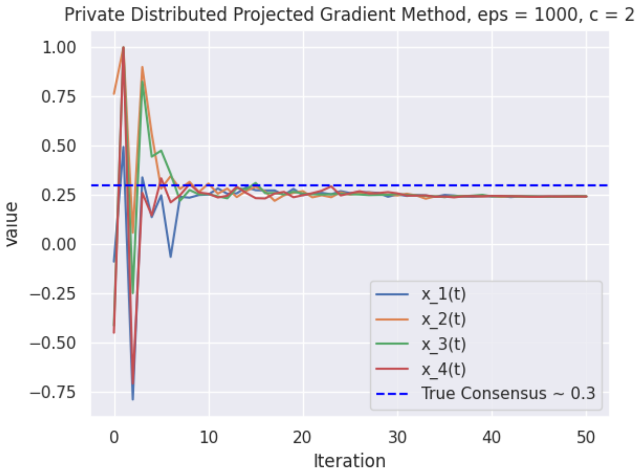

Fig. 4. A suggested relation between ϵ and c such that the system doesn't find the consensus, but still gives insight.

V. Considering more users (no privacy)

Now, considering a pool of 8 users ($n = 8$). With

$$v = \begin{bmatrix} 0.1 & 0.5 & 0.4 & 0.2 & 0.1 & 0.5 & 0.4 & 0.2 \end{bmatrix}$$

We will check with the same variables and model from III if the system still reaches consensus. The new connectivity matrix will be:

$$A = \frac{1}{n}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \text{ For n = 8}$$

We expect to still reach consensus, but slower due to the fact that more people make a system more private. And as can be seen in 5, consensus is indeed still being reached.

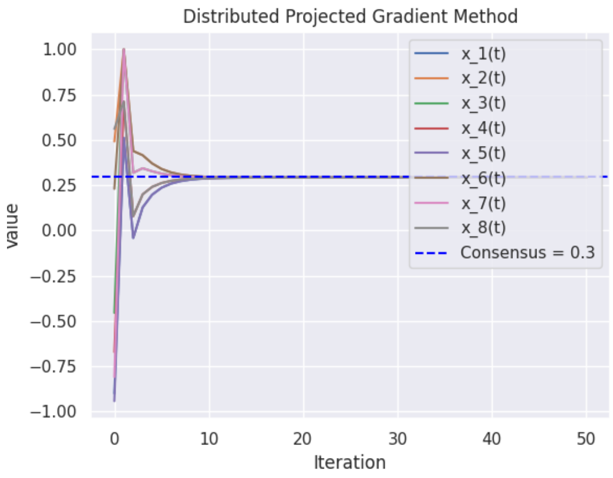

Fig. 5. With n=8, there is still consensus.

VI. Considering more users (with privacy)

Now, the Laplace noise will be added back in with an $\epsilon = 0.1$ and $c = 1$. We expect our data to be moderately private, and as can be seen from 6, this is the case.

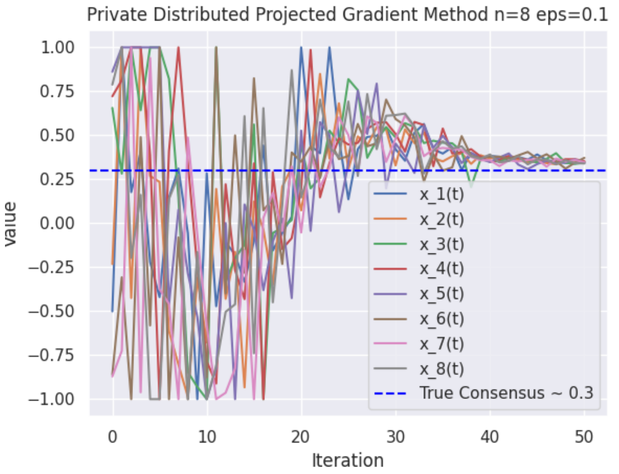

Fig. 6. group of 8 users, data is moderately private

After this ϵ was changed to 10^5. And rerun. Again we expect with such a high value for ϵ, that our system will reach its consensus. And indeed that is correct from looking at 7.

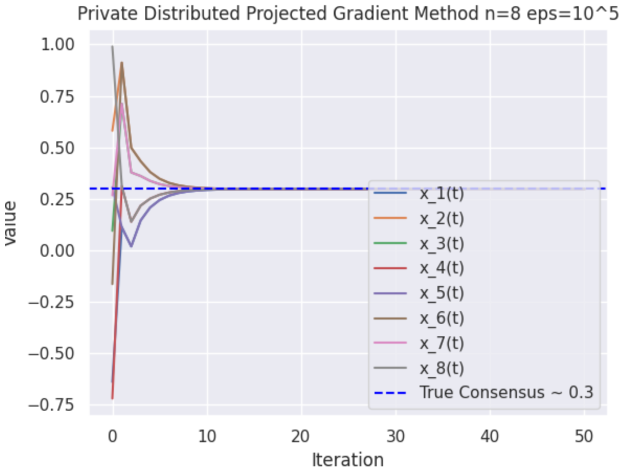

Fig. 7. Reaching consensus for very high ϵ

VII. Conclusion

In this report we ran tests to benchmark the (Private) Distributed Projected Gradient Method using Laplace noise and changing ϵ and c values. Improving insight (enlarging ϵ) will result in higher chances of the users reaching consensus. Which reduces privacy.

On the other hand having a low value for ϵ (e.g 0.1 or 0.001). Will result in a system where no definite answer can be made clear, and the Laplace noise will likely reach its boundary values (as can be seen from 2.

Appendix: Requirements

The formal definition in: ”Privacy in Control and Dynamical Systems” [1] of a connectivity matrix is:

1) Matrix A is nonnegative.
2) The matrix A is doubly stochastic: $\sum_{j=1}^{n} a_{ij} = 1$ for all i and $\sum_{i=1}^{n} a_{ij} = 1$ for all j.
3) The matrix A is irreducible. This implies that the communication graph represented by A is strongly connected.
4) There exists $\eta \in (0, 1]$ such that $a_{ij} \geq \eta$ if user i is able to receive information from user j. As a convention, each user is able to communicate with itself, so that we also have $a_{ii} \geq \eta$ for all i.

References

[1] S. Han and G. Pappas, *Privacy in Control and Dynamical Systems*. Annual Review, 2018.