Robust Combination of Distributed Gradients Under Adversarial Perturbations – Supplemental Document

Kwang In Kim
UNIST

Experiments with random gradient perturbations. In the main paper, we simulated *random labeling* perturbations. Here, we report the results of experiments performed with *random gradient* perturbations which replace the local gradients with independently and identically distributed Gaussian random vectors. Table 1 summarizes the results. The results are consistent with the experiments with the random labeling perturbations. Federated averaging (*FedAvg*) performance rapidly degraded as the number of affected clients increased. The *Median* estimators were comparable to or slightly worse than *FedAvg*. The geometric median estimators (*GeoMed*) were significantly better than *Median*, yet it was noticeably inferior to *FedAvg* in a few cases. Overall, *MedTh* and ours constantly outperformed *FedAvg* and achieved the best performance.

Combining *MedTh* and our algorithm. Our final algorithm estimates the continuous gradient combination weights $\mathbf{a}$. This helps suppress perturbed gradients, but it does not completely rule out their contributions. Here, we demonstrate that we can achieve this goal by combining ours with the thresholding approach of *MedTh*: Each element of $\mathbf{a}$ is set at zero when it is smaller than a threshold. For the experiments with *MedTh*, we determined this threshold such that the number of selected gradients is the same as the number of genuine gradients. As such information is not available in practice, we determine this number using a separate validation set at the server: We performed experiments with a validation set of the same size as the test set. Since the accuracy of the learner network is low at the early stages of training, we exercised this strategy only after 200 iterations of the training process. This (*MedTh + Ours* in Table 2) led to significant performance improvements over our original design. However, as *MedTh + Ours* requires validation sets at the server, its application domain can be limited.

Dataset	Method	Accuracy					
% clients affected		10	20	30	40	50	60
Caltech256	*GTC*	76.54 (0.34)	75.96 (0.20)	75.47 (0.19)	74.84 (0.41)	73.83 (0.16)	72.26 (0.50)
	FedAvg	76.48 (0.32)	75.77 (0.51)	74.81 (0.19)	73.63 (0.24)	72.12 (0.19)	69.61 (0.42)
	Median	72.05 (0.49)	70.13 (0.79)	68.19 (0.52)	64.41 (0.55)	58.92 (1.24)	50.24 (1.60)
	GeoMed	73.17 (0.37)	73.20 (0.31)	73.11 (0.26)	73.05 (0.31)	*73.22 (0.27)*	**73.33 (0.45)**
	MedTh	**76.84 (0.36)**	**75.98 (0.31)**	*75.10 (0.25)*	*74.45 (0.38)*	73.01 (0.40)	71.13 (0.52)
	Ours	*76.63 (0.33)*	*75.89 (0.44)*	**75.23 (0.32)**	**74.50 (0.34)**	**73.58 (0.28)**	*72.25 (0.41)*
CIFAR10	*GTC*	89.53 (0.12)	89.51 (0.18)	89.45 (0.19)	89.53 (0.23)	89.41 (0.14)	89.20 (0.11)
	FedAvg	89.26 (0.15)	*89.21 (0.13)*	88.87 (0.07)	88.42 (0.27)	87.86 (0.16)	87.10 (0.11)
	Median	88.02 (0.15)	87.88 (0.19)	87.34 (0.11)	86.98 (0.17)	86.30 (0.39)	85.47 (0.31)
	GeoMed	87.11 (0.08)	87.22 (0.16)	87.11 (0.08)	87.78 (1.17)	87.09 (0.27)	87.97 (1.34)
	MedTh	*89.27 (0.25)*	89.12 (0.20)	*89.18 (0.15)*	*88.99 (0.20)*	*88.82 (0.16)*	*88.70 (0.12)*
	Ours	**89.33 (0.28)**	**89.47 (0.27)**	**89.40 (0.14)**	**89.37 (0.15)**	**89.27 (0.27)**	**88.95 (0.17)**
CIFAR100	*GTC*	66.86 (0.23)	66.50 (0.21)	66.10 (0.33)	65.81 (0.30)	65.14 (0.42)	64.41 (0.45)
	FedAvg	66.33 (0.26)	65.66 (0.22)	64.68 (0.24)	63.45 (0.47)	61.77 (0.47)	59.33 (0.50)
	Median	66.47 (0.31)	66.25 (0.26)	65.59 (0.23)	65.11 (0.50)	64.62 (0.36)	63.58 (0.32)
	GeoMed	63.09 (0.44)	62.90 (0.33)	62.83 (0.24)	62.76 (0.39)	63.03 (0.26)	62.98 (0.39)
	MedTh	**66.67 (0.31)**	**66.29 (0.19)**	**66.03 (0.43)**	**65.66 (0.39)**	*64.96 (0.29)*	**64.29 (0.37)**
	Ours	*66.66 (0.22)*	*66.27 (0.19)*	*66.00 (0.27)*	*65.55 (0.51)*	**65.02 (0.22)**	*64.17 (0.37)*
FashionMNIST	*GTC*	86.96 (0.18)	86.84 (0.23)	86.84 (0.11)	86.92 (0.19)	86.77 (0.19)	86.77 (0.20)
	FedAvg	86.46 (0.15)	85.83 (0.23)	85.21 (0.24)	84.26 (0.29)	83.32 (0.27)	81.78 (0.38)
	Median	84.26 (0.38)	83.82 (0.32)	83.05 (0.27)	82.07 (0.47)	81.14 (0.36)	80.18 (0.35)
	GeoMed	85.32 (0.91)	85.79 (1.52)	84.74 (0.23)	84.85 (0.19)	85.63 (0.98)	84.76 (0.33)
	MedTh	*86.58 (0.11)*	*86.45 (0.20)*	*86.33 (0.23)*	*86.39 (0.20)*	*86.30 (0.21)*	*86.32 (0.29)*
	Ours	**86.92 (0.18)**	**86.78 (0.20)**	**86.67 (0.23)**	**86.64 (0.13)**	**86.65 (0.11)**	**86.42 (0.21)**
CINIC10	*GTC*	79.69 (0.08)	79.73 (0.14)	79.60 (0.16)	79.67 (0.06)	79.69 (0.13)	79.59 (0.11)
	FedAvg	79.35 (0.13)	*79.20 (0.05)*	78.67 (0.10)	78.03 (0.07)	77.30 (0.17)	75.96 (0.43)
	Median	76.99 (0.06)	76.41 (0.19)	76.04 (0.23)	74.86 (0.25)	74.36 (0.45)	73.18 (0.39)
	GeoMed	77.49 (0.03)	77.72 (0.16)	77.65 (0.32)	77.51 (0.13)	77.54 (0.19)	77.41 (0.08)
	MedTh	*79.45 (0.15)*	79.18 (0.03)	*79.19 (0.25)*	*79.22 (0.06)*	*79.30 (0.08)*	*79.20 (0.07)*
	Ours	**80.02 (0.09)**	**79.95 (0.07)**	**79.81 (0.12)**	**79.88 (0.06)**	**79.84 (0.06)**	**79.65 (0.14)**

Table 1. Results of classification under random gradient perturbations: Mean classification accuracy and standard deviation in parenthesis (both in %; higher is better) on random gradient perturbations. The best and second-best results (except for GTC) are highlighted with **bold** and *italic*, respectively. For *Median*, *MedTh*, and Ours, the results of statistical significance test with respect to *FedAvg* (t–test with $\alpha = 0.95$) are shown: Blue and orange respectively represents significantly better and worse results than *FedAvg*.

Dataset	Method	Accuracy					
% clients affected		10	20	30	40	50	60
Caltech256	*GTC*	76.54 (0.34)	75.96 (0.20)	75.47 (0.19)	74.84 (0.41)	73.83 (0.16)	72.26 (0.50)
	FedAvg	76.29 (0.32)	75.16 (0.35)	74.14 (0.45)	72.29 (0.33)	70.50 (0.27)	66.75 (0.44)
	Median	64.80 (0.47)	63.58 (0.97)	61.82 (0.74)	59.15 (0.98)	55.57 (0.85)	48.15 (0.99)
	GeoMed	76.29 (0.36)	75.68 (0.41)	74.69 (0.44)	71.70 (0.53)	66.49 (0.36)	60.77 (0.82)
	MedTh	76.28 (0.27)	75.59 (0.38)	74.70 (0.85)	72.51 (1.17)	69.74 (1.04)	46.31 (5.19)
	Ours	*76.39 (0.20)*	*75.95 (0.29)*	*75.06 (0.26)*	*73.72 (0.31)*	*71.90 (0.50)*	*69.25 (0.47)*
	MedTh + Ours	**76.76 (0.34)**	**76.01 (0.08)**	**75.56 (0.33)**	**74.88 (0.17)**	**73.93 (0.20)**	**72.48 (0.22)**
CIFAR10	*GTC*	89.57 (0.11)	89.49 (0.15)	89.44 (0.15)	89.55 (0.20)	89.36 (0.16)	89.22 (0.10)
	FedAvg	89.23 (0.10)	87.34 (0.23)	86.58 (0.15)	85.98 (0.37)	85.24 (0.41)	84.83 (0.37)
	Median	87.21 (0.25)	86.89 (0.26)	86.74 (0.82)	86.30 (1.37)	84.65 (1.19)	82.01 (0.62)
	GeoMed	87.38 (0.47)	87.49 (1.06)	87.14 (1.07)	86.89 (1.83)	84.99 (0.85)	82.46 (0.32)
	MedTh	89.39 (0.18)	89.27 (0.10)	89.01 (0.19)	88.57 (0.25)	78.10 (19.90)	31.40 (18.43)
	Ours	**89.51 (0.06)**	*89.43 (0.17)*	*89.22 (0.19)*	*89.14 (0.13)*	*88.95 (0.16)*	*88.85 (0.20)*
	MedTh + Ours	89.47 (0.09)	**89.47 (0.09)**	**89.49 (0.17)**	**89.25 (0.18)**	**89.27 (0.16)**	**89.19 (0.16)**
CIFAR100	*GTC*	66.86 (0.23)	66.50 (0.21)	66.10 (0.33)	65.81 (0.30)	65.14 (0.42)	64.41 (0.45)
	FedAvg	66.50 (0.40)	65.95 (0.54)	64.67 (0.28)	63.25 (0.40)	61.99 (0.32)	60.53 (0.33)
	Median	65.76 (0.23)	64.53 (0.29)	62.73 (0.31)	59.51 (0.75)	53.51 (0.85)	47.89 (0.65)
	GeoMed	**67.11 (0.40)**	**66.80 (0.63)**	65.75 (0.69)	65.20 (0.23)	58.63 (0.57)	49.79 (1.05)
	MedTh	66.80 (0.14)	66.35 (0.29)	65.78 (0.24)	64.11 (0.33)	38.34 (1.43)	36.93 (2.52)
	Ours	66.80 (0.23)	66.41 (0.38)	*65.87 (0.38)*	*65.44 (0.27)*	*64.80 (0.20)*	*63.56 (0.62)*
	MedTh + Ours	66.99 (0.32)	66.46 (0.31)	**65.91 (0.31)**	**65.70 (0.36)**	**65.14 (0.28)**	**64.54 (0.38)**
FashionMNIST	*GTC*	86.96 (0.18)	86.84 (0.23)	86.84 (0.11)	86.92 (0.19)	86.77 (0.19)	86.77 (0.20)
	FedAvg	86.48 (0.19)	83.12 (0.24)	81.93 (0.35)	81.30 (0.31)	80.80 (0.50)	80.47 (0.26)
	Median	83.23 (1.44)	82.77 (1.30)	82.68 (1.71)	82.02 (1.57)	80.44 (1.05)	78.40 (0.44)
	GeoMed	83.29 (1.65)	82.37 (0.40)	82.42 (1.72)	81.20 (0.36)	80.41 (0.45)	78.37 (0.30)
	MedTh	86.75 (0.11)	86.55 (0.22)	86.27 (0.22)	83.57 (4.56)	76.16 (14.07)	52.65 (23.84)
	Ours	**87.01 (0.23)**	*86.73 (0.16)*	*86.63 (0.27)*	*86.40 (0.51)*	*86.24 (0.16)*	*86.41 (0.27)*
	MedTh + Ours	86.92 (0.28)	**86.82 (0.23)**	**86.73 (0.22)**	**86.64 (0.19)**	**86.47 (0.16)**	**86.43 (0.27)**
CINIC10	*GTC*	79.69 (0.07)	79.66 (0.11)	79.65 (0.12)	79.70 (0.08)	79.66 (0.12)	79.60 (0.15)
	FedAvg	78.20 (0.11)	76.88 (0.21)	76.04 (0.16)	75.41 (0.29)	74.81 (0.27)	74.38 (0.13)
	Median	77.16 (1.59)	76.91 (1.61)	77.03 (1.95)	76.84 (1.33)	73.83 (1.44)	71.04 (0.72)
	GeoMed	78.13 (1.95)	79.01 (1.86)	77.49 (2.61)	78.92 (1.20)	75.64 (1.28)	71.51 (0.75)
	MedTh	79.48 (0.11)	79.31 (0.15)	79.06 (0.10)	78.57 (0.15)	57.10 (23.87)	42.19 (20.96)
	Ours	**80.03 (0.13)**	**79.95 (0.17)**	**79.85 (0.16)**	**79.64 (0.32)**	**79.19 (0.50)**	*78.69 (0.52)*
	MedTh + Ours	79.68 (0.09)	79.69 (0.13)	79.25 (0.41)	79.00 (0.60)	79.06 (0.57)	**79.15 (0.42)**

Table 2. Results of classification under random labeling perturbations: Mean accuracy and standard deviation (in parenthesis) are shown in % (higher is better). The best and second-best results excluding *GTC* are highlighted with **bold** and *italic*, respectively. For *Median*, *GeoMed*, *MedTh*, and ours, the results of statistical significance test with respect to *FedAvg* (t–test with $\alpha = 0.95$) are shown: Blue and orange respectively represents significantly better and worse results than *FedAvg*.