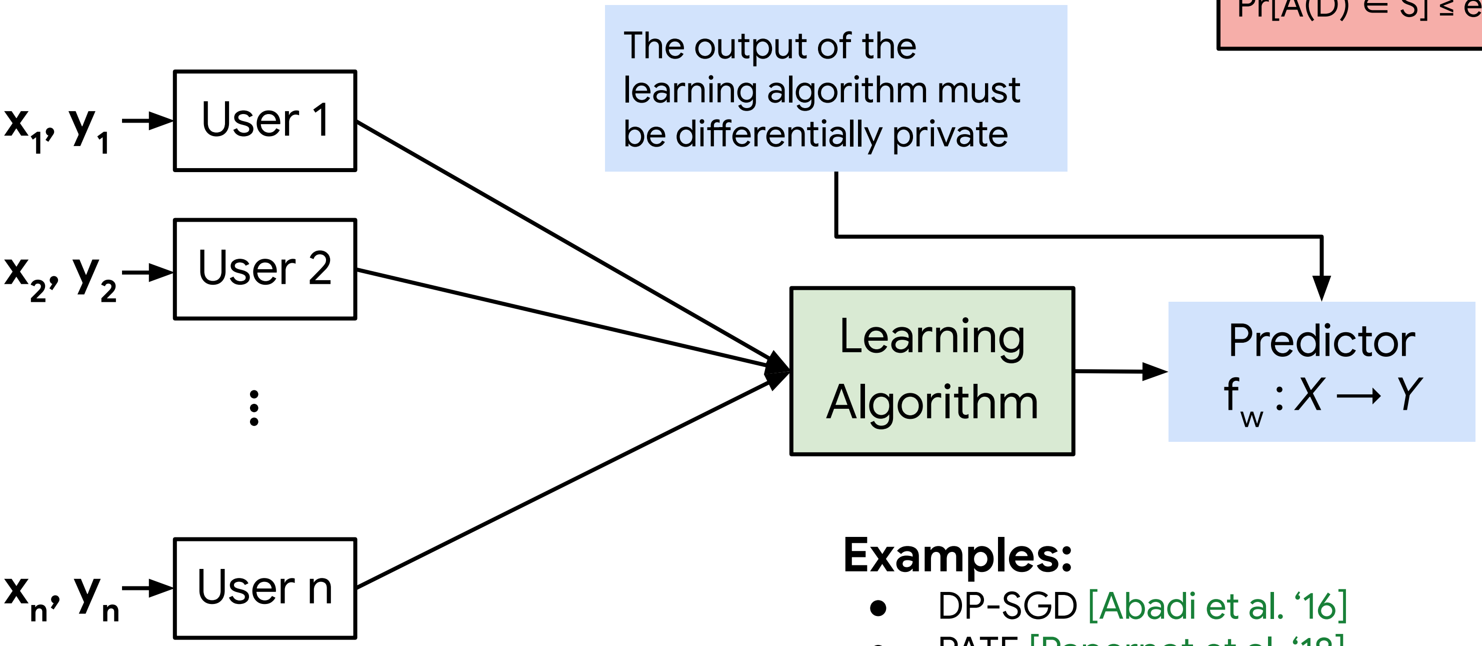
Optimal Unbiased Randomizers for Regression with Label Differential Privacy

Ashwinkumar Badanidiyuru   Badih Ghazi   Pritish Kamath   Ravi Kumar

Ethan Leeman   Pasin Manurangsi   Avinash Varadarajan   Chiyuan Zhang

Differential Privacy [Dwork et al. '06]

(ε, δ)-Differential Privacy
[Dwork et al.'06]
For all S, and two neighboring D, D'
$\Pr[A(D) \in S] \leq e^{\varepsilon} \cdot \Pr[A(D') \in S] + \delta$

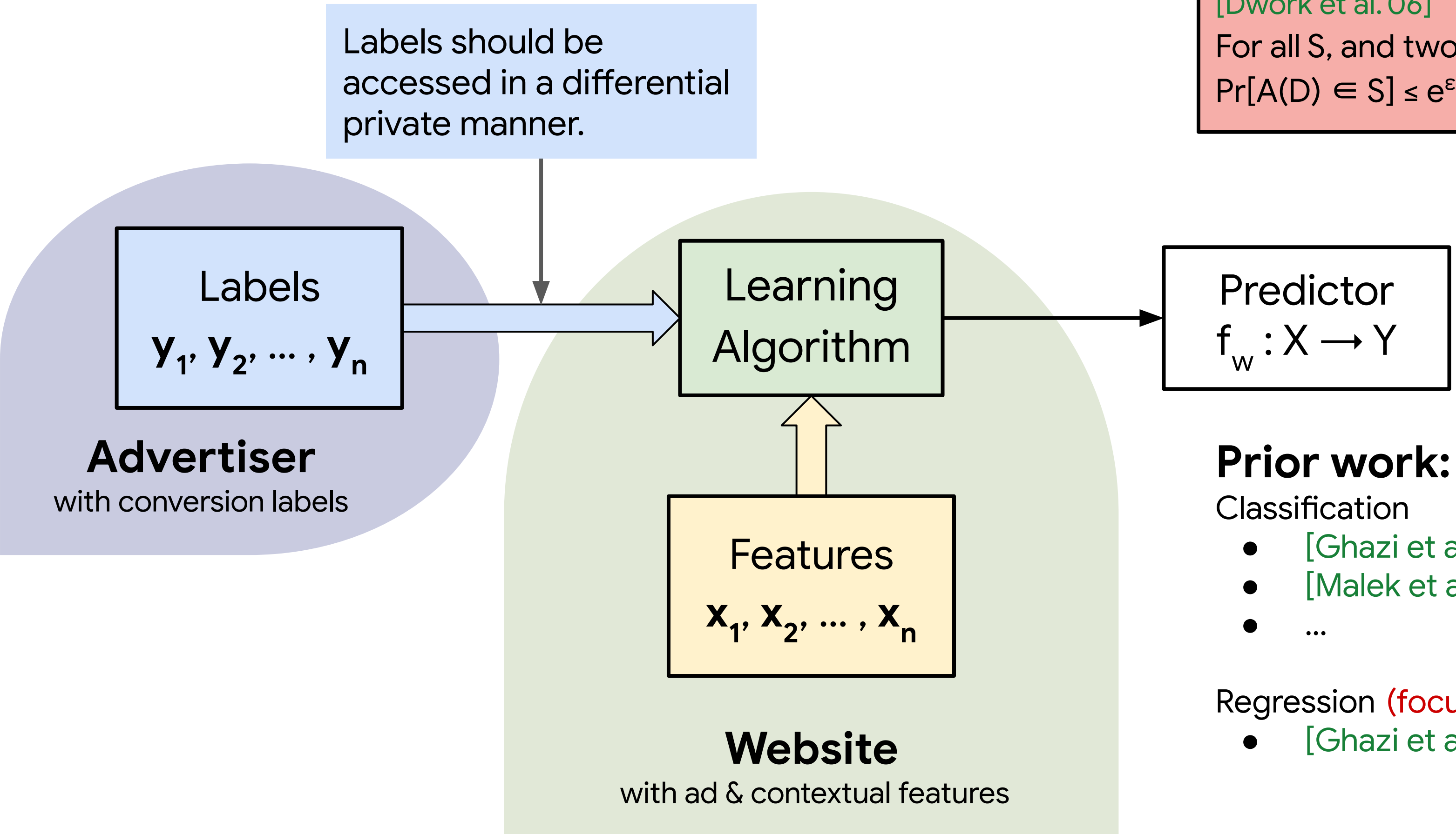


Examples:

- DP-SGD [Abadi et al. '16]
- PATE [Papernot et al. '18]
- DP-FTRL [Kairouz et al. '21]
- ...

Label Differential Privacy [Chaudhuri-Hsu ‘11]

(ε, δ)-Differential Privacy
[Dwork et al.’06]
For all S, and two neighboring D, D’
$\Pr[A(D) \in S] \leq e^{\varepsilon} \cdot \Pr[A(D') \in S] + \delta$

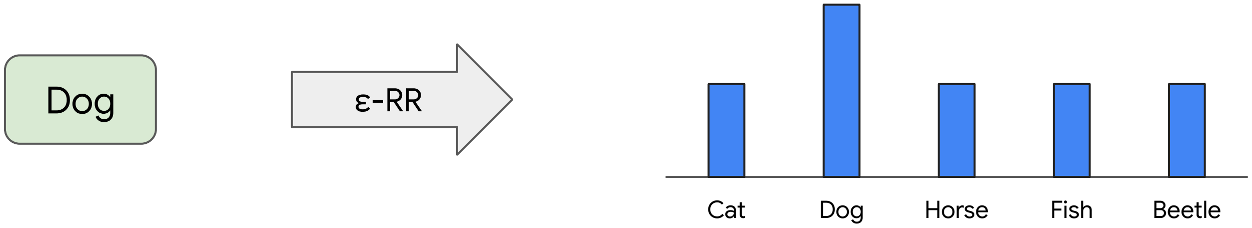


Prior work:

Classification
- [Ghazi et al. ‘21]
- [Malek et al. ‘21]
- ...

Regression (focus of this work)
- [Ghazi et al. ‘23]

Baseline method: Randomized Response



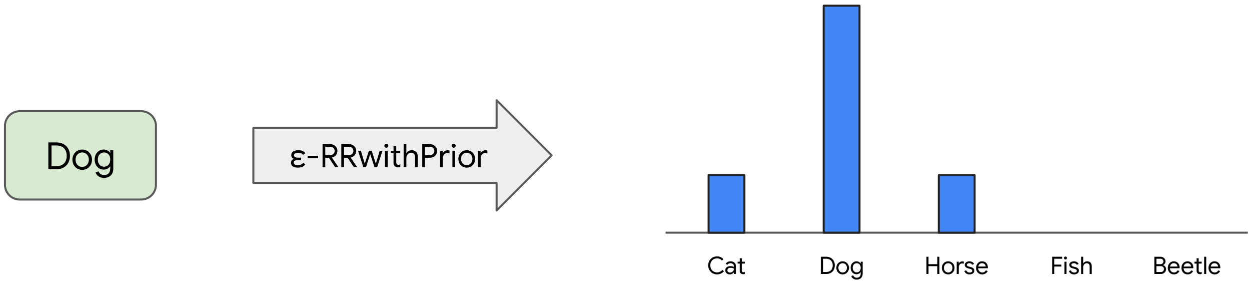
ε-RR

Pr[true label] = e^{ε} / (e^{ε} + k - 1)

Pr[other label] = 1 / (e^{ε} + k - 1)

Essentially outputs random label when ε is small and k is large.

Prior work (Classification): Randomized Response *with Prior*



[Ghazi et al. '21]

Use a *prior P* over labels to choose a better mechanism.

$$\min_{M} \Pr_{\substack{y \sim P \\ y' \sim M(y)}} [y' \neq y]$$

<u>subject to:</u> M is ε-DP.

Mechanisms using prior

Regression [Ghazi et al. ‘23]

$$\min_{M} \mathop{\mathbb{E}}_{\substack{y \sim P \\ y' \sim M(y)}} \left[(y' - y)^2\right]$$

<u>subject to:</u> M is ε-DP.

Classification [Ghazi et al. ‘21]

$$\min_{M} \Pr_{\substack{y \sim P \\ y' \sim M(y)}} [y' \neq y]$$

<u>subject to:</u> M is ε-DP.

Linear program
(for fixed inputs Y, outputs Y’)

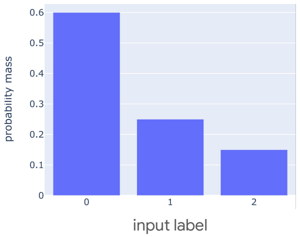
Regression [This work]

$$\min_{M} \mathop{\mathbb{E}}_{\substack{y \sim P \\ y' \sim M(y)}} \left[(y' - y)^2\right]$$

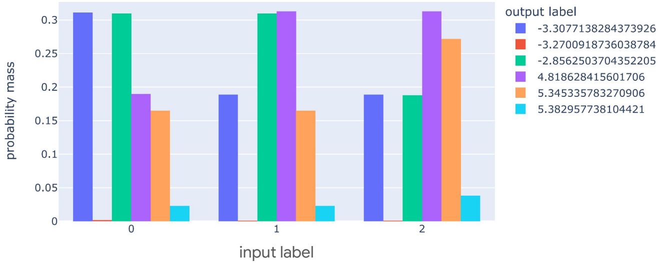
<u>subject to:</u> M is ε-DP

and $\forall y : \mathop{\mathbb{E}}_{y' \sim M(y)} [y'] = y$

Prior P



Optimal Unbiased Mechanism M for ε = 0.5



Loss function: $\ell(\hat{y}, y) := \frac{1}{2}(\hat{y} - y)^2$

Motivation for Unbiased Noisy Labels

Regression [Ghazi et al. '23]

$$\min_{M} \mathop{\mathbb{E}}_{\substack{y \sim P \\ y' \sim M(y)}} \left[(y' - y)^2\right]$$

subject to: M is ε-DP.

Regression [This work]

$$\min_{M} \mathop{\mathbb{E}}_{\substack{y \sim P \\ y' \sim M(y)}} \left[(y' - y)^2\right]$$

subject to: M is ε-DP

and $\forall y : \mathop{\mathbb{E}}_{y' \sim M(y)} [y'] = y$

Minimizing variance, while having zero bias.

- Zero bias preserves the *Bayes Optimal Predictor*

Theorem. The following are equivalent:

- **∀ D :** Predictor minimizing loss w.r.t. **noisy labels** = Predictor minimizing loss w.r.t. **true labels**
- Mechanism is unbiased, that is, $\forall y : \mathop{\mathbb{E}}_{y' \sim M(y)} [y'] = y$

- Zero bias provides unbiased stochastic gradients

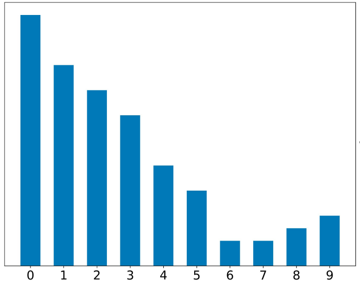
$$\mathop{\mathbb{E}}_{y' \sim M(y)} \nabla_\theta \ell(f_\theta(x), y') = \nabla_\theta \ell(f_\theta(x), y)$$

Since gradient is affine in the label: $\nabla_\theta \ell(f_\theta(x), y) = f_\theta(x) \cdot \nabla_\theta f_\theta(x) - y \cdot \nabla_\theta f_\theta(x)$

Final mechanism: Using privately estimated prior

Prior

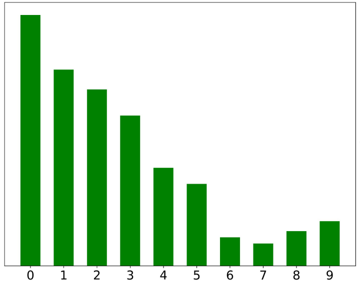
(normalized histogram of labels in training set)



ε_{prior}-DP

Private Prior

(adding Laplace noise to histogram)



$$\min_{M} \mathop{\mathbb{E}}_{\substack{y \sim P \\ y' \sim M(y)}} \left[(y' - y)^2 \right]$$

subject to: M is ε_{label}-DP

and $\forall y \; : \; \mathop{\mathbb{E}}_{y' \sim M(y)} [y'] = y$

M

(Optimal Unbiased Randomizer)

Privacy budget split: $\varepsilon = \varepsilon_{prior} + \varepsilon_{label}$

Choose output set Y' as fine enough grid with heuristic endpoints

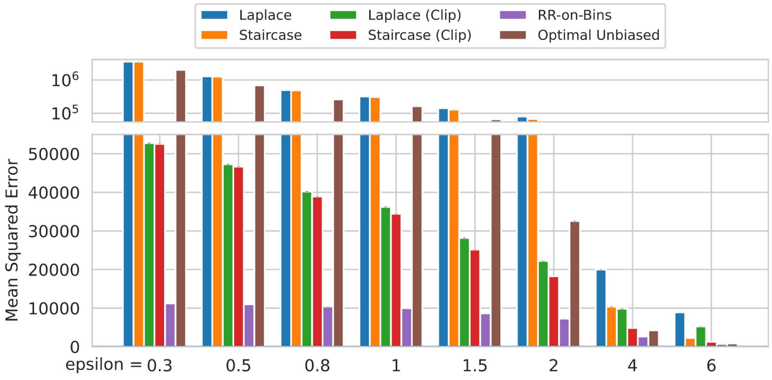
Apply ε_{label}-DP randomizer M to every label in training set.

Evaluation on Criteo Conversion Log Dataset

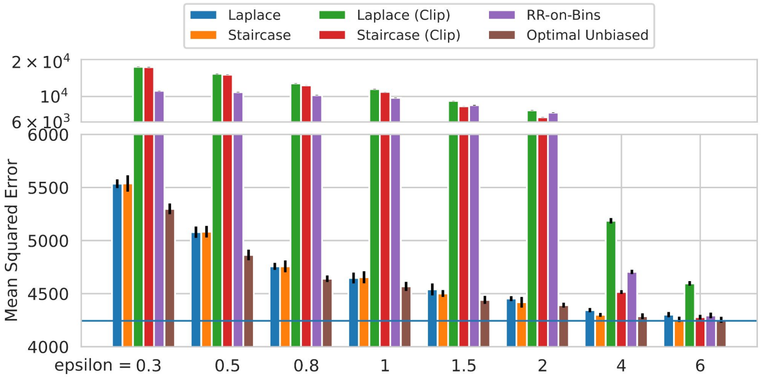
- Criteo Sponsored Search Conversion Log Dataset:
  90 days of Criteo live traffic data, with ~15M examples.
  ailab.criteo.com/criteo-sponsored-search-conversion-log-dataset/

- Goal: Predict conversion value (in €)
  (clipped to €400 for simplicity)

Noisy label loss on train data: $\frac{1}{n_{\text{train}}}\sum_{i=1}^{n_{\text{train}}}(y_i - y_i')^2$



Prediction loss on test data: $\frac{1}{n_{\text{test}}}\sum_{i=1}^{n_{\text{test}}}(f_w(x_i) - y_i)^2$



Badanidiyuru, Ghazi, Kamath, Kumar, Leeman, Manurangsi, Varadarajan, Zhang

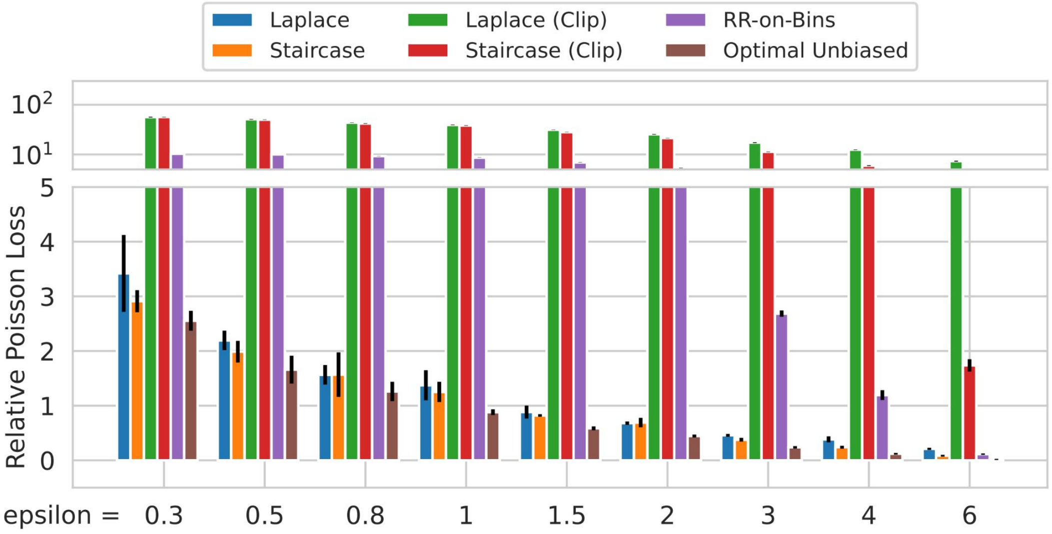
Evaluation on App Ads Conversion Count

- Commercial mobile app store conversion count prediction dataset.
- Goal: predict the number of post-click conversion events in the app after user install within timeframe.

Plot shows relative Poisson log loss compared to non-private baseline.

Poisson log loss:

$$\frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} f_w(x_i) - y_i \log(f_w(x_i))$$



Highlights

- Unbiased randomizers eliminate bias, at the cost of increased variance.
- Optimal unbiased randomizers provide most utility.
- Partial characterization of optimal unbiased randomizers:
  - “Staircase mechanism” [Kairouz et al. ‘16]
  - Bound on number of output labels.

Future Directions

- Full characterization of optimal unbiased randomizers?
- Better algorithms for computing optimal unbiased randomizers?

Thanks!

Mechanisms using prior

Classification [Ghazi et al. '21]

$$\min_{M} \Pr_{\substack{y \sim P \\ y' \sim M(y)}} [y' \neq y]$$

subject to: M is ε-DP.

Regression [Ghazi et al. '23]

$$\min_{M} \mathop{\mathbb{E}}_{\substack{y \sim P \\ y' \sim M(y)}} [\ell(y', y)]$$

subject to: M is ε-DP.

Examples:

- $\ell(y', y) = (y' - y)^2$
- $\ell(y', y) = |y' - y|$

Linear program

(for fixed inputs Y, outputs Y')

Regression [This work]

$$\min_{M} \mathop{\mathbb{E}}_{\substack{y \sim P \\ y' \sim M(y)}} [\ell(y', y)]$$

subject to: M is ε-DP

and $\forall y : \mathop{\mathbb{E}}_{y' \sim M(y)} [y'] = y$

Bias-Variance Trade-off

Regression [Ghazi et al. '23]

$$\min_{M} \mathop{\mathbb{E}}_{\substack{y \sim P \\ y' \sim M(y)}} [\ell(y', y)]$$

subject to: M is ε-DP.

Regression [This work]

$$\min_{M} \mathop{\mathbb{E}}_{\substack{y \sim P \\ y' \sim M(y)}} [\ell(y', y)]$$

subject to: M is ε-DP

and $\forall y : \mathop{\mathbb{E}}_{y' \sim M(y)} [y'] = y$

Bias–Variance Trade-offs

Training Set: $S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$

Training Set with randomized labels: $S' = \{(x_1, y'_1), \ldots, (x_n, y'_n)\}$

Loss function: $\ell(\hat{y}, y) := \frac{1}{2}(\hat{y} - y)^2$

Regression [Ghazi et al. '23]

$$\min_{M} \mathop{\mathbb{E}}_{\substack{y \sim P \\ y' \sim M(y)}} \left[(y' - y)^2\right]$$

<u>subject to:</u> M is ε-DP.

Regression [This work]

$$\min_{M} \mathop{\mathbb{E}}_{\substack{y \sim P \\ y' \sim M(y)}} \left[(y' - y)^2\right]$$

<u>subject to:</u> M is ε-DP

and $\forall y : \mathop{\mathbb{E}}_{y' \sim M(y)} [y'] = y$

Batch gradient using noisy labels — Population gradient

$$\nabla_\theta \mathcal{L}_{S'}(f_\theta) - \nabla_\theta \mathcal{L}_{\mathcal{D}}(f_\theta)$$

$$\|$$

$$\underbrace{\nabla_\theta \mathcal{L}_S(f_\theta) - \nabla_\theta \mathcal{L}_{\mathcal{D}}(f_\theta)}_{\text{Statistical error}} + \underbrace{\mathbb{E}_{(x,y)\in S}(y - \mathbb{E}\, y') \cdot \nabla_\theta f_\theta(x) + \mathbb{E}_{(x,y)\in S}(\mathbb{E}\, y' - y') \cdot \nabla_\theta f_\theta(x)}_{\text{Error due to privacy}}$$

Minimizing variance, subject to having zero bias.

$$\nabla_\theta \ell(f_\theta(x), y) = (f_\theta(x) - y) \cdot \nabla_\theta f_\theta(x)$$

$$\mathop{\mathbb{E}}_{\substack{(x,y) \sim \mathcal{D} \\ y' \sim M(y)}} [\ell(f_\theta(x), y')]$$