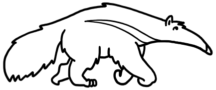
Randomized Edge-Assisted On-Sensor Information Selection for Bandwidth-Constrained Systems

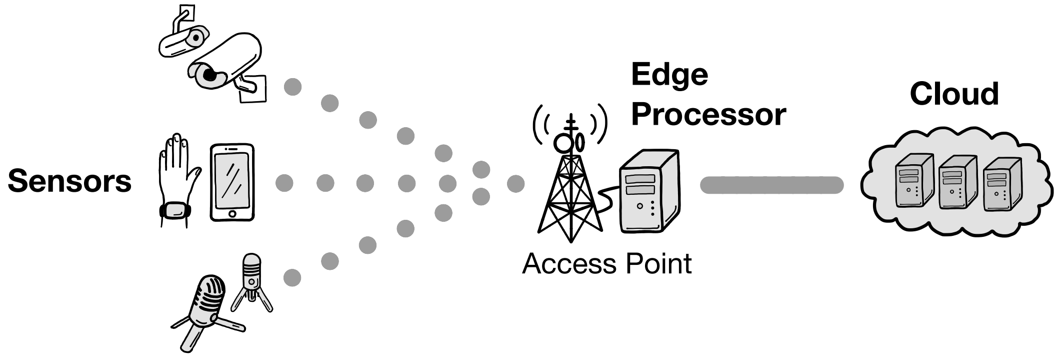
Igor Burago   Marco Levorato

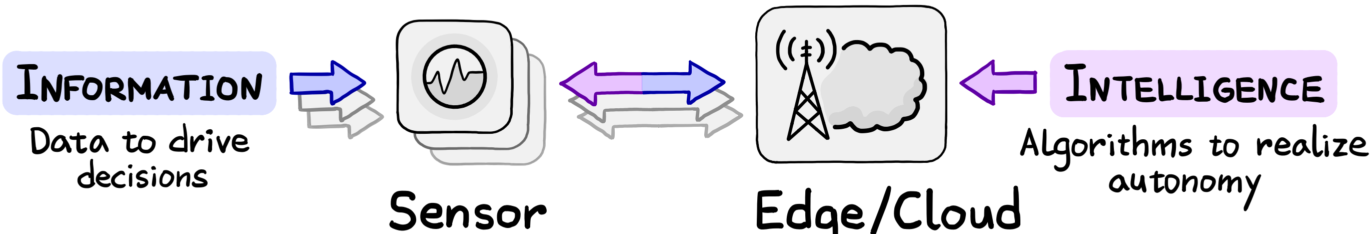
Department of Computer Science
University of California, Irvine



2018

Intelligent IoT in Principle





Sensor

- Observes the environment
- Selectively informs the edge
- Adjusts its own operation

Edge/Cloud

- Accumulates relevant observations
- Exercises high-level control
- Assists on-sensor decision-making

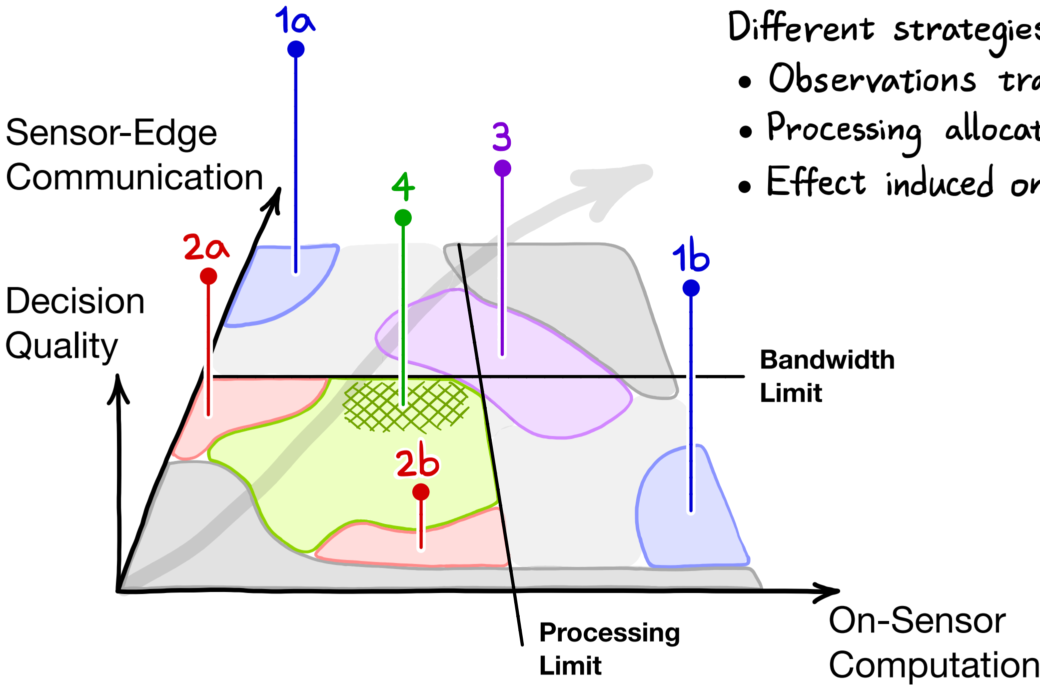
Goal: Realize the alliance between information completeness and best intelligence

Intelligent IoT in Principle

Computation-communication tradeoffs

Different strategies vary in the amount of:

- Observations transferred to the edge
- Processing allocated on the sensor
- Effect induced on the quality of decisions



1a Offload all decisions to edge

1b Offload all decisions to sensor

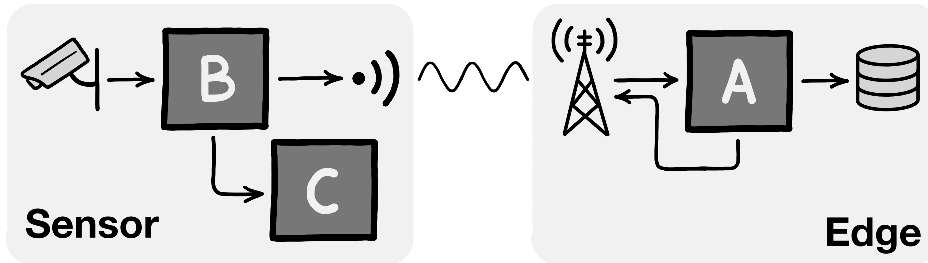
3 Transmit all in compressed form

2a Sample observations to transmit

2b Transmit obvious positives

4 Optimal adaptive compromise

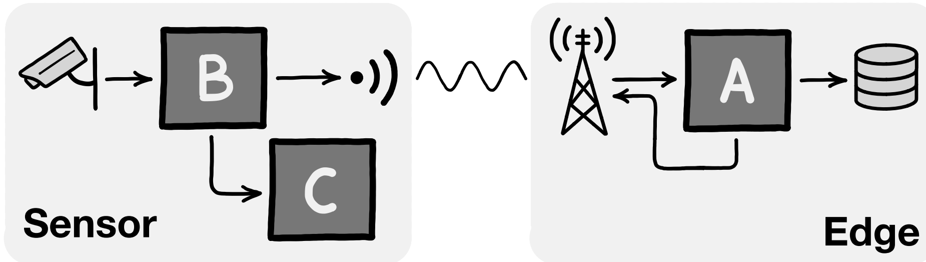
Sensor-Edge Cooperation Model



Model assumptions

- Observations fall into two classes:
  - of interest to the system (positives)
  - and not (negatives)
- Positives are to be collected at edge for further processing in the system
- Edge can run the optimal classifier δ; sensor can not
- $\boxed{B}$ Sensor runs a low-complexity classifier supplied by edge
- $\boxed{A}$ Edge, via supervision of its classifier, trains the sensor's
- $\boxed{C}$ Sensor uses its classifier to spare bandwidth by dropping some negatives, and to make local decisions

Sensor-Edge Cooperation Model

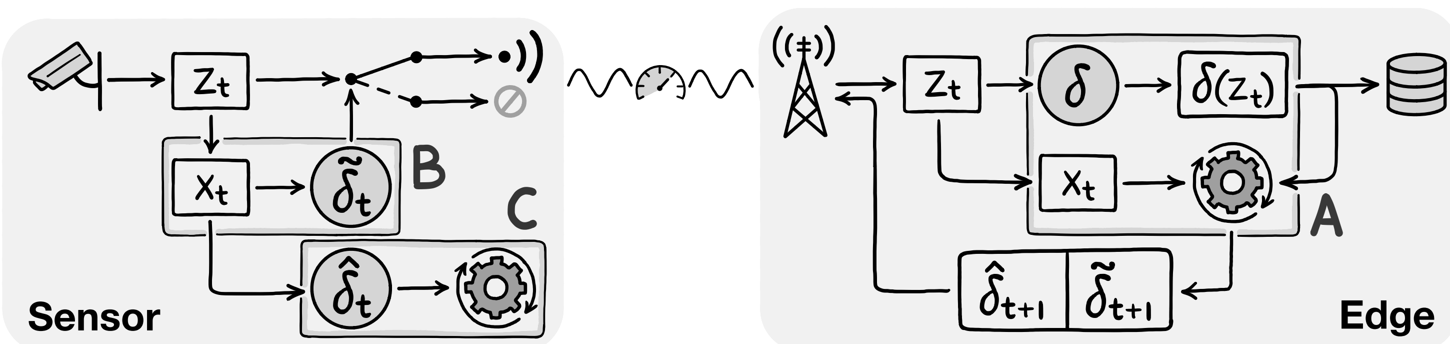


Factors at play

- CPU, battery, etc.
- Computation constraints
- Complexity of $\tilde{\delta}$ and $\hat{\delta}$

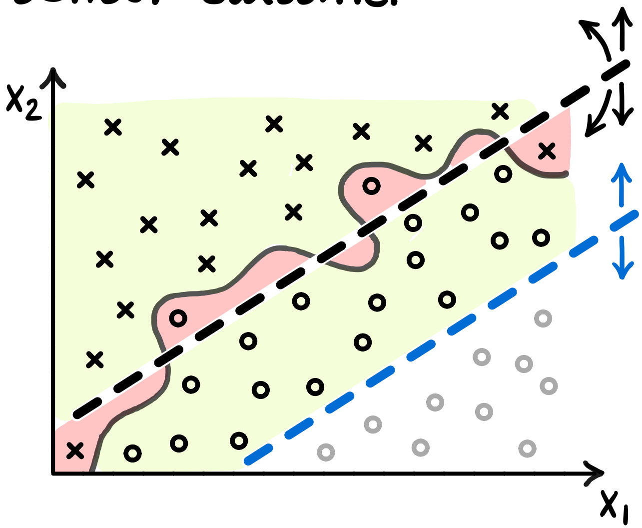
- Shared wireless channel
- Bandwidth constraint
- Transmission probability $\varphi \approx \varphi_*$

- Need to react to environment
- Cost of error in feature space
- Expected risk U to minimize



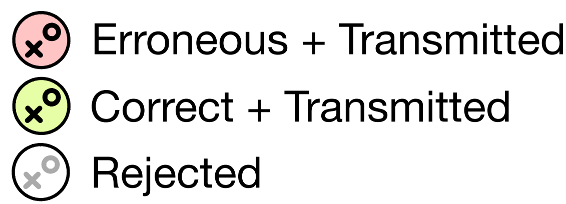
Sensor-Edge Cooperation Model

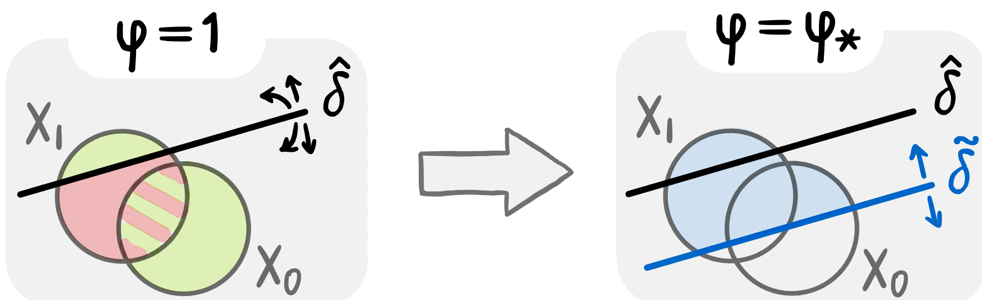
On-sensor classifier



$$\hat{\delta}: f(x,\theta) \triangleq \frac{\langle x,\theta\rangle}{\|\theta\|} \gtrless \eta_1$$

$$\tilde{\delta}: f(x,\theta) \gtrless \eta_0$$





Sensor-Edge Cooperation Model

Problem

$$U(\tau) \longrightarrow \min, \quad \text{s.t. } W(\tau) = 0 \quad \Rightarrow \quad \frac{1}{\zeta} U(\tau) + W(\tau) \longrightarrow \min_{\tau}$$

$$\underbrace{U(\tau) \triangleq \mathbb{E}\left[\hat{I}_{\eta_0}(u_\theta)\, I_e(z,\tau)\, \left| r_0(u_\theta - \eta_0)\, r_1(u_\theta - \eta_1) \right| \right]}_{\text{Expected penalty for errors in transmitted data}} \qquad \underbrace{W(\tau) \triangleq \frac{1}{2}\left(\varphi(\tau) - \varphi_*\right)^2}_{\text{Constraint penalty}}$$

$$\underbrace{\tau = \mathrm{vec}[\theta, \eta_0, \eta_1]}_{\text{Parameters}} \qquad \underbrace{u_\theta \triangleq f(x,\theta)}_{\text{Decision function}} \qquad \underbrace{\eta_0,\ \eta_1}_{\text{Thresholds}} \qquad \underbrace{\varphi(\tau) \triangleq \mathbb{E}[\hat{I}_{\eta_0}(u_\theta)]}_{\text{Transmission probability}}$$

$\hat{I}_{\eta_0}$, $\hat{I}_{\eta_1}$, $I_1 \in \{0,1\}$ indicate whether an observation z with features x is to be transmitted, classified as a positive, and is a true positive.

$I_e(z,\tau) \triangleq (1 - I_1(z))\,\hat{I}_{\eta_1}(u_\theta) + I_1(z)\left(1 - \hat{I}_{\eta_1}(u_\theta)\right)$ is the error indicator.

r_0, r_1 are loss functions based on "distances" $u_\theta - \eta_0$ and $u_\theta - \eta_1$.
In the simplest case, $r_i(u_\theta - \eta_i) = u_\theta - \eta_i$.

Decision Rules Update

Stochastic quasi-gradient algorithm

$$\theta_{t+1} = \theta_t - \gamma_{1,t}\,\hat{G}_\theta(\bar{z}_{t+1}, \tau_t) - \gamma_{2,t}\,\hat{g}_\theta(\tilde{x}^\varphi_{t+1}, \tilde{x}^\psi_{t+1}, \tau_t, \varepsilon_t)$$
$$\eta_{0,t+1} = \eta_{0,t} - \gamma_{3,t}\,\hat{g}_{\eta_0}(\tilde{x}^\varphi_{t+1}, \tilde{x}^\psi_{t+1}, \tau_t, \varepsilon_t)$$
$$\eta_{1,t+1} = \max\left\{\eta_{0,t+1},\ \eta_{1,t} - \gamma_{1,t}\,\hat{G}_{\eta_1}(\bar{z}_{t+1}, \tau_t)\right\} \qquad \text{(a)}$$

Quasi-gradient of risk U

$$\hat{G}_\theta(\{z_i\}_{i=1}^n, \tau) \triangleq \frac{1}{n}\sum_{i=1}^n \left[J(z_i, \tau)\,\nabla_\theta\left[r_0(u_i - \eta_0)\,r_1(u_i - \eta_1)\right]\right]$$
$$\hat{G}_{\eta_1}(\{z_i\}_{i=1}^n, \tau) \triangleq -\frac{1}{n}\sum_{i=1}^n \left[J(z_i, \tau)\,r_0(u_i - \eta_0)\,\frac{d}{du}r_1(u_i - \eta_1)\right]$$
where $x_i = \chi(z_i)$, $u_i = f(x_i, \theta)$,
$$J(z, \tau) \triangleq \hat{I}_{\eta_0}(f(\chi(z), \theta))\left(\hat{I}_{\eta_1}(f(\chi(z), \theta)) - I_1(z)\right) \qquad \text{(b)}$$

Quasi-gradient of constraint

$$\hat{g}_\theta(\tilde{x}^\varphi, \tilde{x}^\psi, \tau, \varepsilon) \triangleq \left(\hat{\varphi}(\tilde{x}^\varphi, \tau) - \varphi_*\right)\hat{\psi}_\theta(\tilde{x}^\psi, \tau, \varepsilon)$$
$$\hat{g}_{\eta_0}(\tilde{x}^\varphi, \tilde{x}^\psi, \tau, \varepsilon) \triangleq \left(\hat{\varphi}(\tilde{x}^\varphi, \tau) - \varphi_*\right)\hat{\psi}_\eta(\tilde{x}^\psi, \tau, \varepsilon)$$
$$\hat{\varphi}(\{x_i\}_{i=1}^n, \tau) \triangleq \frac{1}{n}\sum_{i=1}^n \hat{I}_{\eta_0}(f(x_i, \theta))$$
$$\hat{\psi}_\theta(\{x_i\}_{i=1}^n, \tau, \varepsilon) \triangleq \frac{1}{\varepsilon n}\sum_{i=1}^n \left[I_{S(\tau,\varepsilon)}(f(x_i, \theta))\,\nabla_\theta f(x_i, \theta)\right]$$
$$\hat{\psi}_\eta(\{x_i\}_{i=1}^n, \tau, \varepsilon) \triangleq -\hat{\psi}_0 = \text{const} < 0 \qquad \text{(c)}$$

Decision Rules Update

$$\theta_{t+1} = \theta_t - \gamma_{1,t}\hat{G}_{\theta}(\bar{z}_{t+1}, \tau_t) - \gamma_{2,t}\hat{g}_{\theta}(\tilde{x}^{\psi}_{t+1}, \tilde{x}^{\psi}_{t+1}, \tau_t, \varepsilon_t)$$

$$\eta_{0,t+1} = \eta_{0,t} - \gamma_{3,t}\hat{g}_{\eta_0}(\tilde{x}^{\psi}_{t+1}, \tilde{x}^{\psi}_{t+1}, \tau_t, \varepsilon_t)$$

$$\eta_{1,t+1} = \max\{\eta_{0,t+1}, \ \eta_{1,t} - \gamma_{1,t}\hat{G}_{\eta_1}(\bar{z}_{t+1}, \tau_t)\}$$

a

a Stochastic quasi-gradient algorithm

- b Quasi-gradient $\hat{G}$ of the risk U
- c Quasi-gradient $\hat{g}$ of the constraint term W

Decision Rules Update

$$\theta_{t+1} = \theta_t - \gamma_{1,-} \hat{G}_\theta(\bar{z}_{t+1}, \tau_t) - \gamma_{2,t}\, \hat{g}_\theta(\tilde{x}^\psi_{t+1}, \tilde{x}^\psi_{t+1}, \tau_t, \varepsilon_t)$$

$$\eta_{0,t+1} = \eta_{0,t} - \gamma_{3,t}\, \hat{g}_{\eta_0}(\tilde{x}^\psi_{t+1}, \tilde{x}^\psi_{t+1}, \tau_t, \varepsilon_t)$$

$$\eta_{1,t+1} = \max\{\eta_{0,t+1},\ \eta_{1,t} - \gamma_{1,t}\, \hat{G}_{\eta_1}(\bar{z}_{t+1}, \tau_t)\}$$

a

a Stochastic quasi-gradient algorithm

- b Quasi-gradient $\hat{G}$ of the risk U
- c Quasi-gradient $\hat{g}$ of the constraint term W
- Updates are done in batches $\bar{z}_{t+1} = \{z_i\}$, with instrumental sub-batches $\tilde{x}^\psi_{t+1}, \tilde{x}^\psi_{t+1}$ extracted from $\bar{z}_{t+1}$.

Decision Rules Update

$$\theta_{t+1} = \theta_t - \gamma_{1,t}\,\widehat{G}_\theta(\bar{z}_{t+1}, \tau_t) - \gamma_{2,t}\,\hat{g}_\theta(\tilde{x}^{\varphi}_{t+1}, \tilde{x}^{\psi}_{t+1}, \tau_t, \varepsilon_t)$$
$$\eta_{0,t+1} = \eta_{0,t} - \gamma_{3,t}\,\hat{g}_{\eta_0}(\tilde{x}^{\varphi}_{t+1}, \tilde{x}^{\psi}_{t+1}, \tau_t, \varepsilon_t)$$
$$\eta_{1,t+1} = \max\{\eta_{0,t+1},\ \eta_{1,t} - \gamma_{1,t}\,\widehat{G}_{\eta_1}(\bar{z}_{t+1}, \tau_t)\} \quad (a)$$

(a) Stochastic quasi-gradient algorithm

- (b) Quasi-gradient $\widehat{G}$ of the risk U
- (c) Quasi-gradient $\hat{g}$ of the constraint term W
- Updates are done in batches $\bar{z}_{t+1} = \{z_i\}$, with instrumental sub-batches $\tilde{x}^{\varphi}_{t+1}$, $\tilde{x}^{\psi}_{t+1}$ extracted from $\bar{z}_{t+1}$.
- Threshold η_1 is clipped with η_0 for the bandwidth constraint to prevail over the classification risk.

Decision Rules Update

$$\theta_{t+1} = \theta_t - \gamma_{1,t}\hat{G}_\theta(\bar{z}_{t+1}, \tau_t) - \gamma_{2,t}\hat{g}_\theta(\tilde{x}^{\varphi}_{t+1}, \tilde{x}^{\psi}_{t+1}, \tau_t, \varepsilon_t)$$
$$\eta_{0,t+1} = \eta_{0,t} - \gamma_{3,t}\hat{g}_{\eta_0}(\tilde{x}^{\varphi}_{t+1}, \tilde{x}^{\psi}_{t+1}, \tau_t, \varepsilon_t)$$
$$\eta_{1,t+1} = \max\{\eta_{0,t+1},\ \eta_{1,t} - \gamma_{1,t}\hat{G}_{\eta_1}(\bar{z}_{t+1}, \tau_t)\} \quad \text{(a)}$$

(a) Stochastic quasi-gradient algorithm

- (b) Quasi-gradient $\hat{G}$ of the risk U
- (c) Quasi-gradient $\hat{g}$ of the constraint term W
- Updates are done in batches $\bar{z}_{t+1} = \{z_i\}$, with instrumental sub-batches $\tilde{x}^{\varphi}_{t+1}$, $\tilde{x}^{\psi}_{t+1}$ extracted from $\bar{z}_{t+1}$.
- Threshold η_1 is clipped with η_0 for the bandwidth constraint to prevail over the classification risk.
- Different learning rates $\gamma_{1,t}, \gamma_{2,t}, \gamma_{3,t}$.

Decision Rules Update

$$\theta_{t+1} = \theta_t - \gamma_{1,t}\,\hat{G}_\theta(\bar{z}_{t+1}, \tau_t) - \gamma_{2,t}\,\hat{g}_\theta(\tilde{x}^{\varphi}_{t+1}, \tilde{x}^{\psi}_{t+1}, \tau_t, \varepsilon_t)$$
$$\eta_{0,t+1} = \eta_{0,t} - \gamma_{3,t}\,\hat{g}_{\eta_0}(\tilde{x}^{\varphi}_{t+1}, \tilde{x}^{\psi}_{t+1}, \tau_t, \varepsilon_t)$$
$$\eta_{1,t+1} = \max\left\{\eta_{0,t+1},\ \eta_{1,t} - \gamma_{1,t}\,\hat{G}_{\eta_1}(\bar{z}_{t+1}, \tau_t)\right\} \quad \text{(a)}$$

$$\hat{G}_\theta(\{z_i\}_{i=1}^n, \tau) \triangleq \frac{1}{n}\sum_{i=1}^{n}\Big[J(z_i,\tau)\,\nabla_\theta\big[r_0(u_i-\eta_0)\,r_1(u_i-\eta_1)\big]\Big]$$
$$\hat{G}_{\eta_1}(\{z_i\}_{i=1}^n, \tau) \triangleq -\frac{1}{n}\sum_{i=1}^{n}\Big[J(z_i,\tau)\,r_0(u_i-\eta_0)\,\frac{d}{du}r_1(u_i-\eta_1)\Big]$$
where $x_i = \chi(z_i)$, $u_i = f(x_i,\theta)$,
$$J(z,\tau) \triangleq \hat{I}_{\eta_0}(f(\chi(z),\theta))\,\big(\hat{I}_{\eta_1}(f(\chi(z),\theta)) - I_1(z)\big) \quad \text{(b)}$$

(b) Quasi-gradient $\hat{G}$ of the risk U

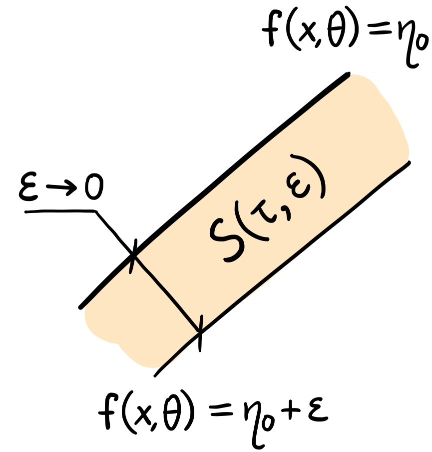
$\mathbb{E}[\hat{G}(\bar{z},\tau)] = G(\tau)$, the generalized gradient such that

$U(\tau_{t+1}) - U(\tau_t) \leq \langle \tau_{t+1}-\tau_t, G(\tau_t)\rangle + O(\|\tau_{t+1}-\tau_t\|^2)$.

Decision Rules Update

$$\hat{g}_\theta(\tilde{x}^\varphi, \tilde{x}^\psi, \tau, \varepsilon) \triangleq (\hat{\varphi}(\tilde{x}^\varphi, \tau) - \varphi_*)\,\hat{\psi}_\theta(\tilde{x}^\psi, \tau, \varepsilon)$$
$$\hat{g}_{\eta_0}(\tilde{x}^\varphi, \tilde{x}^\psi, \tau, \varepsilon) \triangleq (\hat{\varphi}(\tilde{x}^\varphi, \tau) - \varphi_*)\,\hat{\psi}_\eta(\tilde{x}^\psi, \tau, \varepsilon)$$
$$\hat{\varphi}(\{x_i\}_{i=1}^n, \tau) \triangleq \frac{1}{n}\sum_{i=1}^n \hat{I}_{\eta_0}(f(x_i, \theta))$$
$$\hat{\psi}_\theta(\{x_i\}_{i=1}^n, \tau, \varepsilon) \triangleq \frac{1}{\varepsilon n}\sum_{i=1}^n \left[I_{S(\tau,\varepsilon)}(f(x_i,\theta))\,\nabla_\theta f(x_i,\theta)\right]$$
$$\hat{\psi}_\eta(\{x_i\}_{i=1}^n, \tau, \varepsilon) \triangleq -\hat{\psi}_0 = \text{const} < 0$$

c



c Quasi-gradient $\hat{g}$ of the constraint term W

- Evolution of the integral in the gradient of W

$\nabla_\tau W(\tau) = (\varphi(\tau) - \varphi_*)\,\psi(\tau)$, where

$$\psi(\tau) \triangleq \nabla_\tau \mathbb{E}[\hat{I}_{\eta_0}(u_\theta)] = \int_{\{f(x,\theta)=\eta_0\}} \frac{\nabla_\tau[f(x,\theta)-\eta_0]}{\|\nabla_x f(x,\theta)\|}\,p(x)\,d\sigma$$
$$= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{S(\tau,\varepsilon)} \nabla_\tau[f(x,\theta)-\eta_0]\,p(x)\,dx$$

volume

↓

surface

↓

volume

$$S(\tau,\varepsilon) \triangleq \{x : \eta_0 < f(x,\theta) \leq \eta_0 + \varepsilon\}, \qquad I_{S(\tau,\varepsilon)}(u) \triangleq \mathbb{1}[u \in S(\tau,\varepsilon)]$$

Decision Rules Update

$$\theta_{t+1} = \theta_t - \gamma_{1,t}\hat{G}_\theta(\bar{z}_{t+1}, \tau_t) - \gamma_{2,t}\hat{g}_\theta(\tilde{x}^\varphi_{t+1}, \tilde{x}^\psi_{t+1}, \tau_t, \varepsilon_t)$$

$$\eta_{0,t+1} = \eta_{0,t} - \gamma_{3,t}\hat{g}_{\eta_0}(\tilde{x}^\varphi_{t+1}, \tilde{x}^\psi_{t+1}, \tau_t, \varepsilon_t)$$

$$\eta_{1,t+1} = \max\{\eta_{0,t+1},\ \eta_{1,t} - \gamma_{1,t}\hat{G}_{\eta_1}(\bar{z}_{t+1}, \tau_t)\}$$

a

$$\hat{g}_\theta(\tilde{x}^\varphi, \tilde{x}^\psi, \tau, \varepsilon) \triangleq (\hat{\varphi}(\tilde{x}^\varphi, \tau) - \varphi_*)\,\hat{\psi}_\theta(\tilde{x}^\psi, \tau, \varepsilon)$$

$$\hat{g}_{\eta_0}(\tilde{x}^\varphi, \tilde{x}^\psi, \tau, \varepsilon) \triangleq (\hat{\varphi}(\tilde{x}^\varphi, \tau) - \varphi_*)\,\hat{\psi}_\eta(\tilde{x}^\psi, \tau, \varepsilon)$$

$$\hat{\varphi}(\{x_i\}_{i=1}^n, \tau) \triangleq \frac{1}{n}\sum_{i=1}^n \hat{I}_{\eta_0}(f(x_i, \theta))$$

$$\hat{\psi}_\theta(\{x_i\}_{i=1}^n, \tau, \varepsilon) \triangleq \frac{1}{\varepsilon n}\sum_{i=1}^n \left[I_{S(\tau,\varepsilon)}(f(x_i, \theta))\,\nabla_\theta f(x_i, \theta)\right]$$

$$\hat{\psi}_\eta(\{x_i\}_{i=1}^n, \tau, \varepsilon) \triangleq -\hat{\varphi}_0 = \text{const} < 0$$

c

- Instrumental samples are used to make a stochastic estimator for the product of φ and ψ in $\nabla_\tau W(\tau)$:

$$\mathbb{E}[\hat{g}(\tilde{x}^\varphi, \tilde{x}^\psi, \tau, \varepsilon)] = \mathbb{E}[(\hat{\varphi}(x, \tau) - \varphi_*)\,\hat{\psi}(x, \tau, \varepsilon)] + O(\varepsilon)$$
$$= \mathbb{E}[\hat{\varphi}(\tilde{x}^\varphi, \tau) - \varphi_*]\,\mathbb{E}[\hat{\psi}(\tilde{x}^\psi, \tau, \varepsilon)] + O(\varepsilon)$$

Decision Rules Update

$$\theta_{t+1} = \theta_t - \gamma_{1,t}\,\hat{G}_\theta(\bar{z}_{t+1}, \tau_t) - \gamma_{2,t}\,\hat{g}_\theta(\tilde{x}^\varphi_{t+1}, \tilde{x}^\psi_{t+1}, \tau_t, \varepsilon_t)$$
$$\eta_{0,t+1} = \eta_{0,t} - \gamma_{3,t}\,\hat{g}_{\eta_0}(\tilde{x}^\varphi_{t+1}, \tilde{x}^\psi_{t+1}, \tau_t, \varepsilon_t)$$
$$\eta_{1,t+1} = \max\{\eta_{0,t+1},\ \eta_{1,t} - \gamma_{1,t}\,\hat{G}_{\eta_1}(\bar{z}_{t+1}, \tau_t)\} \quad \text{(a)}$$

$$\hat{g}_\theta(\tilde{x}^\varphi, \tilde{x}^\psi, \tau, \varepsilon) \triangleq (\hat{\varphi}(\tilde{x}^\varphi, \tau) - \varphi_*)\,\hat{\psi}_\theta(\tilde{x}^\psi, \tau, \varepsilon)$$
$$\hat{g}_{\eta_0}(\tilde{x}^\varphi, \tilde{x}^\psi, \tau, \varepsilon) \triangleq (\hat{\varphi}(\tilde{x}^\varphi, \tau) - \varphi_*)\,\hat{\psi}_\eta(\tilde{x}^\psi, \tau, \varepsilon)$$
$$\hat{\varphi}(\{x_i\}_{i=1}^n, \tau) \triangleq \frac{1}{n}\sum_{i=1}^n \hat{I}_{\eta_0}(f(x_i, \theta))$$
$$\hat{\psi}_\theta(\{x_i\}_{i=1}^n, \tau, \varepsilon) \triangleq \frac{1}{\varepsilon n}\sum_{i=1}^n \left[I_{S(\tau,\varepsilon)}(f(x_i, \theta))\,\nabla_\theta f(x_i, \theta)\right]$$
$$\hat{\psi}_\eta(\{x_i\}_{i=1}^n, \tau, \varepsilon) \triangleq -\hat{\psi}_0 = \text{const} < 0 \quad \text{(c)}$$

- As in the above, the gradient of the integral leads to:

$$\hat{\psi}_\eta(\{x_i\}_{i=1}^n, \tau, \varepsilon) \triangleq -\frac{1}{\varepsilon n}\sum_{i=1}^n I_{S(\tau,\varepsilon)}(f(x_i, \theta)) \leq 0$$

Yet, in the algorithm, we use a constant estimator for promptness in the updates of η_0, as higher adaptivity is of more value than accuracy for adjusting to the bandwidth constraint.

Decision Rules Update

Theorem (Algorithm convergence) Let

- batches $\overline{z}_t$ be i.i.d.;
- sub-batches $\tilde{x}^{\varphi}_t$ and $\tilde{x}^{\psi}_t$ be mutually independent;
- $r_0(u)$, $r_1(u)$, and $f(x,\theta)$ be continuously differentiable;
- $\operatorname{sgn}(r_i(u)) = \operatorname{sgn}(u)$, $i \in \{0,1\}$; $r_1(u) \leqslant r_1(v)$ for all $0 \leqslant u < v$;
- p.d.f. $p(x)$ of observations in the feature space be continuous and have compact support;
- $\gamma_{1,t}/\gamma_{2,t}$ and ε_t be monotonically decreasing
  $\sum \gamma^2_{i,t} < \infty$, $i \in \{1,2,3\}$; $\sum \gamma_{j,t}\,\varepsilon_t < \infty$, $j \in \{1,2\}$;
  $\sum \gamma_{3,t}\,\gamma_{1,t}/\gamma_{2,t} < \infty$; $\sum \gamma_{3,t} = \infty$; $\sum \gamma^2_{1,t}/\gamma_{2,t} = \infty$.

Then, for $\zeta_t = \gamma_{2,t}/\gamma_{1,t}$,

- $\frac{1}{\zeta_t} U(\tau_t) + W(\tau_t) \xrightarrow[t\to\infty]{a.s.} V_*$, $\mathbb{E}[V_*] < \infty$;
- $\underline{\lim}_{t\to\infty} |\varphi(\tau_t) - \varphi_*| = 0$;
- $\underline{\lim}_{t\to\infty} G_{\eta_1}(\tau_t) = 0$;
- $\underline{\lim}_{t\to\infty} G_\theta(\tau_t) + \zeta_t \nabla_\theta W(\tau_t) = 0$.