# Online Convex Optimization

## Chapter 2: Complete Treatment with Proofs

Shai Shalev-Shwartz

# Convexification Techniques

## Two Key Methods

- Randomization: Allow probabilistic predictions
- Surrogate Loss: Replace original loss with convex upper bound

## Example (Prediction with Expert Advice)

- Original problem: Discrete choices
- Convexification: $w_t \in \Delta_d$ (probability simplex)
- Loss becomes $\langle w_t, z_t \rangle$

## Example (Online Classification)

- Surrogate loss: $f_t(w) = 2|\langle w, v_t \rangle - y_t|$
- Maintains $w_t \in \Delta_{|H|}$ (version space probabilities)

# FTL Algorithm and Analysis

## FTL Update Rule

$w_t =_{w \in S} \sum_{i=1}^{t-1} f_i(w)$

## Lemma (Regret Decomposition)

*For any* $u \in S$*:* $\mathit{Regret}_T(u) \leq \sum_{t=1}^{T} [f_t(w_t) - f_t(w_{t+1})]$

## Inductive Proof Sketch.

Base case: $T = 1$ trivial. Inductive step uses:
$\sum_{t=1}^{T} f_t(w_t) \leq \sum_{t=1}^{T} f_t(w_{T+1})$ □

## Example (Quadratic Loss)

For $f_t(w) = \frac{1}{2}\|w - z_t\|^2$:

- FTL: $w_t = \frac{1}{t-1} \sum_{i=1}^{t-1} z_i$
- Regret $O(\log T)$

# FTRL Framework

## Regularized Objective

$w_t =_{w \in S} \left( \sum_{i=1}^{t-1} f_i(w) + R(w) \right)$

## Theorem (FTRL Regret Bound)

*For linear* $f_t$ *and* $R(w) = \frac{1}{2\eta}\|w\|^2$*:* $Regret_T(u) \leq \frac{\|u\|^2}{2\eta} + \eta \sum_{t=1}^{T} \|z_t\|^2$

## Key Steps.

- Apply FTL analysis to regularized losses
- Use Lemma 2.3: $\sum(f_t(w_t) - f_t(u)) \leq R(u) + \sum(f_t(w_t) - f_t(w_{t+1}))$
- Bound stability terms via strong convexity

$\square$

## Doubling Trick

Adaptive $\eta$ selection without knowing $T$: $\eta_m = \frac{B}{L\sqrt{2^m}}$ for epoch $m$

# OGD Algorithm and Analysis

**Algorithm 1** Online Gradient Descent

**Require:** $\eta > 0$

Initialize $w_1 = 0$
**for** $t = 1$ to $T$ **do**
    Predict $w_t$, receive $z_t \in \partial f_t(w_t)$
    Update $w_{t+1} = w_t - \eta z_t$
**end for**

## Theorem (OGD Regret Bound)

*For $L$-Lipschitz losses:* $\mathrm{Regret}_T(u) \le \frac{\|u\|^2}{2\eta} + \eta T L^2$

## Proof.

- Use FTRL analysis with $R(w) = \frac{1}{2\eta}\|w\|^2$
- Show $\sum \|z_t\|^2 \le TL^2$

# Strongly Convex Regularizers

**Definition ($\sigma$-Strong Convexity)**

$R(u) \geq R(w) + \langle \nabla R(w), u - w \rangle + \frac{\sigma}{2}\|u - w\|^2$

**Example (Common Regularizers)**

- Euclidean: $R(w) = \frac{1}{2}\|w\|_2^2$ (1-strongly convex)
- Entropic: $R(w) = \sum w_i \log w_i$ (1-strongly convex w.r.t. $\ell_1$)

**Lemma (Implication for FTRL)**

*For $\sigma$-strongly convex $R$:* $\mathrm{Regret}_T(u) \leq \frac{R(u)}{\eta} + \eta \sum_{t=1}^{T} \|z_t\|_*^2$

# OMD and Duality

## OMD Framework

- Primal update: $w_{t+1} =_w \langle \eta z_t, w \rangle + D_R(w \| w_t)$
- Dual view: $\theta_{t+1} = \theta_t - \eta z_t$ with $w_t = \nabla R^*(\theta_t)$

## Theorem (General Regret Bound)

*For $R$ $(1/\eta)$-strongly convex:* $\mathit{Regret}_T(u) \leq R(u) + \eta \sum_{t=1}^{T} \|z_t\|_*^2$

## Example (EG Algorithm)

- Regularizer: $R(w) = \sum w_i \log w_i$
- Update: $w_{t+1,i} \propto w_{t,i} e^{-\eta z_{t,i}}$

# Local Norm Analysis

## Theorem (Normalized EG)

*For* $0 \le z_{t,i} \le 1$*:* $\mathit{Regret}_T(u) \le \frac{\log d}{\eta} + \eta \sum_{t=1}^{T} \sum_i w_{t,i} z_{t,i}^2$

## Key Inequality.

Using $e^{-a} \le 1 - a + a^2$ for $a \ge -1$:

$$D_{R^*}(-z_{1:t} \| - z_{1:t-1}) \le \eta \sum_i w_{t,i} z_{t,i}^2$$

□

## Optimal Tuning

Set $\eta = \sqrt{\frac{\log d}{T}}$ for $O(\sqrt{T \log d})$ regret

# Key References

- Zinkevich (2003): Original OGD framework
- Hazan et al. (2007): Adaptive gradient methods
- Shalev-Shwartz (2011): Survey synthesis
- Rakhlin (2014): Duality approaches
- Cesa-Bianchi & Lugosi (2006): Prediction with expert advice