# Lecture III: Variational Inequalities

ALINA ENE

ADFOCS '21: Convex Optimization and Graph Algorithms

# Problem Definition

$F\colon \mathbb{R}^d \to \mathbb{R}^d$ vector-valued function, called an operator

$K \subseteq \mathbb{R}^d$ convex and bounded constraint set

# Problem Definition

$F: \mathbb{R}^d \to \mathbb{R}^d$ vector-valued function, called an operator

$K \subseteq \mathbb{R}^d$ convex and bounded constraint set

The variational inequality problem associated with $F$ asks for a strong solution, i.e., a point $x^* \in K$ satisfying:

$$\langle F(x^*), x^* - y \rangle \leq 0 \quad \forall y \in K$$

# Problem Definition

$F: \mathbb{R}^d \to \mathbb{R}^d$ vector-valued function, called an operator

$K \subseteq \mathbb{R}^d$ convex and bounded constraint set

The variational inequality problem associated with $F$ asks for a strong solution, i.e., a point $x^* \in K$ satisfying:

$$\langle F(x^*), x^* - y \rangle \leq 0 \quad \forall y \in K$$

A weak solution is a point $x^* \in K$ satisfying

$$\langle F(y), x^* - y \rangle \leq 0 \quad \forall y \in K$$

# Problem Definition

$F\colon \mathbb{R}^d \to \mathbb{R}^d$ vector-valued function, called an operator

$K \subseteq \mathbb{R}^d$ convex and bounded constraint set

The variational inequality problem associated with $F$ asks for a strong solution, i.e., a point $x^* \in K$ satisfying:

$$\langle F(x^*), x^* - y \rangle \leq 0 \quad \forall y \in K$$

A weak solution is a point $x^* \in K$ satisfying

$$\langle F(y), x^* - y \rangle \leq 0 \quad \forall y \in K$$

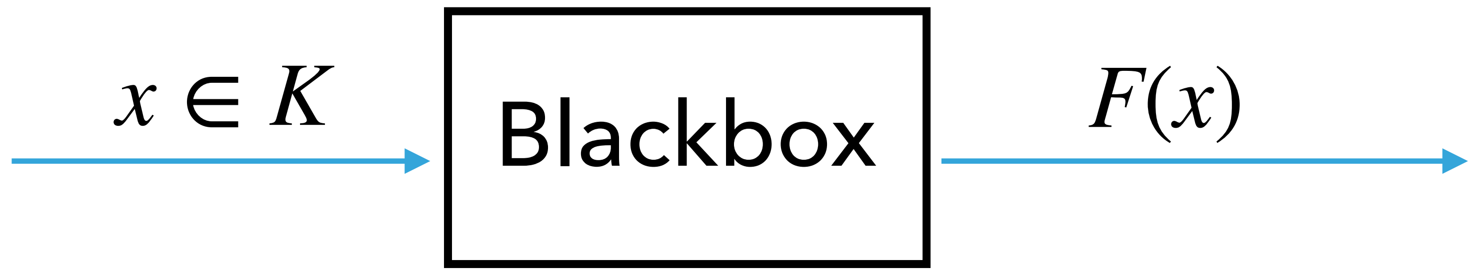
For the operators we will consider in this lecture, a weak solution is a strong solution and vice versa

# Problem Definition

$F: \mathbb{R}^d \to \mathbb{R}^d$ vector-valued function, called an operator

$K \subseteq \mathbb{R}^d$ convex and bounded constraint set

**Computational model:** operator access via ~~first-order~~ operator oracle




**Goal:** minimize number of queries $x_1, x_2, \ldots, x_T$ to obtain

$$\mathrm{Err}(x_{out}) := \sup_{y \in K} \langle F(y), x_{out} - y \rangle \leq \epsilon$$

# Example: Convex Minimization

Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a differentiable loss function

# Example: Convex Minimization

Let $f \colon \mathbb{R}^d \to \mathbb{R}$ be a differentiable loss function

Consider the operator $F(x) = \nabla f(x)$

# Example: Convex Minimization

Let $f: \mathbb{R}^d \to \mathbb{R}$ be a differentiable loss function

Consider the operator $F(x) = \nabla f(x)$

The strong solutions are the minimizers of $f$ over $K$:

$$\langle \nabla f(x^*), x^* - y \rangle \leq 0 \;\; \forall y \in K \Leftrightarrow x^* \in \arg\min_{x \in K} f(x)$$

If $f$ is convex, we have the following $\forall x, y$:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle$$

$$\Rightarrow \quad \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0 \quad \forall x, y$$

This gives us a way to extend convexity to operators

# Example: Convex Minimization

Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a differentiable loss function

Consider the operator $F(x) = \nabla f(x)$

The strong solutions are the minimizers of $f$ over $K$:

$$\langle \nabla f(x^*), x^* - y \rangle \leq 0 \ \ \forall y \in K \Leftrightarrow x^* \in \arg\min_{x \in K} f(x)$$

The operator analogue of convexity is monotonicity:

$$\langle F(x) - F(y), x - y \rangle \geq 0 \quad \forall x, y \in K$$

# Example: Convex Minimization

Let $f: \mathbb{R}^d \to \mathbb{R}$ be a differentiable loss function

Consider the operator $F(x) = \nabla f(x)$

The strong solutions are the minimizers of $f$ over $K$:

$$\langle \nabla f(x^*), x^* - y \rangle \leq 0 \quad \forall y \in K \Leftrightarrow x^* \in \arg\min_{x \in K} f(x)$$

The operator analogue of convexity is monotonicity:

$$\langle F(x) - F(y), x - y \rangle \geq 0 \quad \forall x, y \in K$$

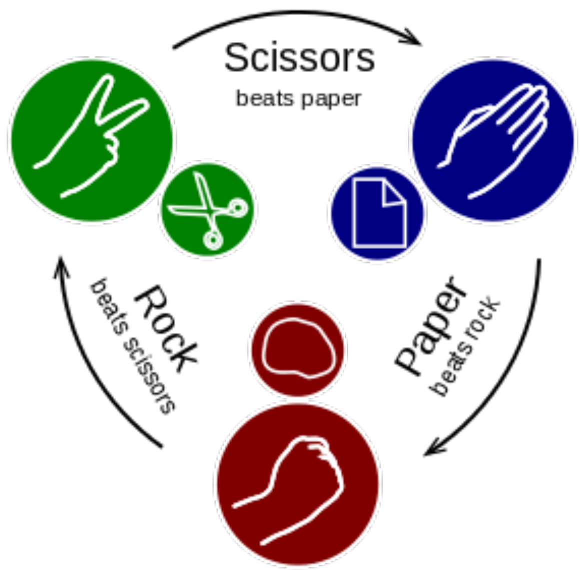
Throughout, we will assume that $F$ is monotone and continuous

For such operators, weak solutions are strong solutions, and we can measure convergence via the error function:

$$\mathrm{Err}(x) := \sup_{y \in K} \langle F(y), x - y \rangle$$

# Example: Nash Equilibria in Games

Consider a 2-player zero-sum game, such as:

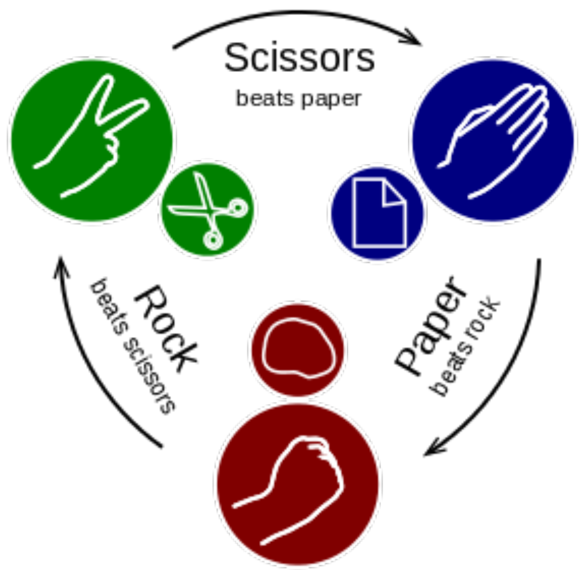



Payoff matrix:

$$\mathbb{A} = \begin{array}{c} R \\ P \\ S \end{array} \overset{\begin{array}{ccc} R & P & S \end{array}}{\begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}}$$

# Example: Nash Equilibria in Games

Consider a 2-player zero-sum game, such as:




Payoff matrix:

$$\mathbb{A} = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}$$

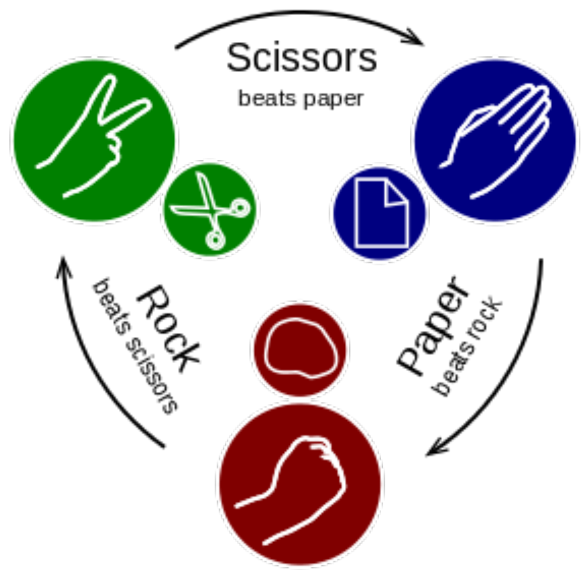
Alice chooses a distribution $p \in \Delta_3$ over the strategies

Bob chooses a distribution $q \in \Delta_3$ over the strategies

Alice's expected payoff is $f(p, q) := p^\top \mathbb{A} q$

# Example: Nash Equilibria in Games

Consider a 2-player zero-sum game, such as:




Payoff matrix:

$$\mathbb{A} = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}$$

Alice chooses a distribution $p \in \Delta_3$ over the strategies

Bob chooses a distribution $q \in \Delta_3$ over the strategies

Alice's expected payoff is $f(p, q) := p^\top \mathbb{A} q$

A pair of strategies $(p^*, q^*)$ is a mixed Nash equilibrium if

$$f(p, q^*) \leq f(p^*, q^*) \leq f(p^*, q) \quad \forall p, q \in \Delta_3$$

No player is better off switching if the other player's strategy remains fixed

# Example: Nash Equilibria in Games

Consider a 2-player zero-sum game with payoff matrix $\mathbb{A} \in \mathbb{R}^{m \times n}$

Alice's expected payoff is $f(p,q) := p^\top \mathbb{A} q$

A pair of strategies $(p^*, q^*)$ is a mixed Nash equilibrium if

$$\max_{p \in \Delta_m} f(p, q^*) \le f(p^*, q^*) \le \min_{q \in \Delta_n} f(p^*, q)$$

Consider the monotone operator:

$$F((p,q)) = \Big(-\nabla_p f(p,q), \nabla_q f(p,q)\Big) = (-\mathbb{A}q, \mathbb{A}^\top p)$$

Suppose $(p^*, q^*)$ is a strong solution.

$$\langle (-Aq^*, A^\top p^*), (p^*, q^*) - (p, q^*) \rangle \le 0 \quad \forall p$$

$$\Rightarrow \quad p^{*\top} A q^* \ge p^\top A q^* \quad \forall p$$

Similarly, $p^{*\top} A q^* \le p^{*\top} A q \quad \forall q$

# Example: Nash Equilibria in Games

Consider a 2-player zero-sum game with payoff matrix $\mathbb{A} \in \mathbb{R}^{m \times n}$

Alice's expected payoff is $f(p, q) := p^\top \mathbb{A} q$

A pair of strategies $(p^*, q^*)$ is a mixed Nash equilibrium if

$$\max_{p \in \Delta_m} f(p, q^*) \leq f(p^*, q^*) \leq \min_{q \in \Delta_n} f(p^*, q)$$

Consider the monotone operator:

$$F((p, q)) = \left( -\nabla_p f(p, q), \nabla_q f(p, q) \right) = (-\mathbb{A} q, \mathbb{A} p)$$

A strong solution for the VI is a Nash equilibrium for the game

# Example: Min-Max Optimization

2-player games are a special case of min-max optimization:

Minimax Thm:
(Von Neumann)

$$\min_{q \in \Delta_n} \max_{p \in \Delta_m} p^\top \mathbb{A} q = \max_{p \in \Delta_m} \min_{q \in \Delta_n} p^\top \mathbb{A} q$$

Suppose players take turns and both play optimally.

If Alice goes first:

Alice's payoff $= \min_{q} \max_{p} p^T A q$

If Alice goes second:

Alice's payoff $= \max_{p} \min_{q} p^T A q$

Minimax Thm: there is no advantage to going second

# Example: Min-Max Optimization

2-player games are a special case of min-max optimization:

$$\min_{q \in \Delta_n} \max_{p \in \Delta_m} p^\top \mathbb{A} q = \max_{p \in \Delta_m} \min_{q \in \Delta_n} p^\top \mathbb{A} q$$

More generally, we can consider the min-max optimization

$$\min_{u \in U} \max_{v \in V} f(u, v)$$

where $f(u, v)$ is convex in $u$ and concave in $v$

# Example: Min-Max Optimization

2-player games are a special case of min-max optimization:

$$\min_{q\in\Delta_n}\max_{p\in\Delta_m} p^\top \mathbb{A} q = \max_{p\in\Delta_m}\min_{q\in\Delta_n} p^\top \mathbb{A} q$$

More generally, we can consider the min-max optimization

$$\min_{u\in U}\max_{v\in V} f(u, v)$$

where $f(u, v)$ is convex in $u$ and concave in $v$

As before, we can consider the monotone operator:

$$F((u, v)) = \left(\nabla_u f(u, v), -\nabla_v f(u, v)\right)$$

# Example: Min-Max Optimization

2-player games are a special case of min-max optimization:

$$\min_{q \in \Delta_n} \max_{p \in \Delta_m} p^\top \mathbb{A} q = \max_{p \in \Delta_m} \min_{q \in \Delta_n} p^\top \mathbb{A} q$$

More generally, we can consider the min-max optimization

$$\min_{u \in U} \max_{v \in V} f(u, v)$$

where $f(u, v)$ is convex in $u$ and concave in $v$

As before, we can consider the monotone operator:

$$F((u, v)) = \left( \nabla_u f(u, v), -\nabla_v f(u, v) \right)$$

or saddle point

A strong solution $(u^*, v^*)$ for the VI is an equilibrium:

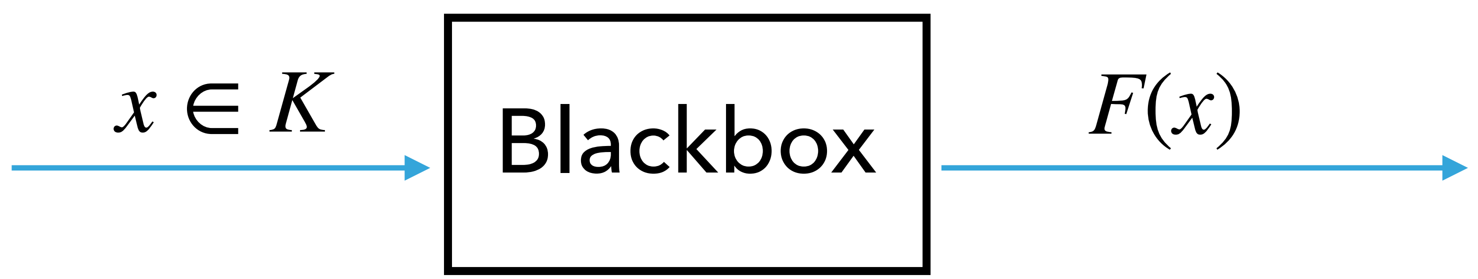
$$\max_{v \in V} f(u^*, v) \leq f(u^*, v^*) \leq \min_{u \in U} f(u, v^*)$$

# Variational Inequalities

$F: \mathbb{R}^d \to \mathbb{R}^d$ monotone operator

$K \subseteq \mathbb{R}^d$ convex and bounded constraint set

**Computational model:** operator access via ~~first-order~~ operator oracle




**Goal:** minimize number of queries $x_1, x_2, \ldots, x_T$ to obtain

$$\mathrm{Err}(x_{out}) := \sup_{y \in K} \langle F(y), x_{out} - y \rangle \le \epsilon$$

# In Gradient Descent We Trust

"Gradient" descent:

$$x_t = \arg\min_{x \in K} \left\{ \langle F(x_{t-1}), x \rangle + \frac{1}{2\eta} \|x - x_{t-1}\|^2 \right\}$$

# In Gradient Descent We Trust

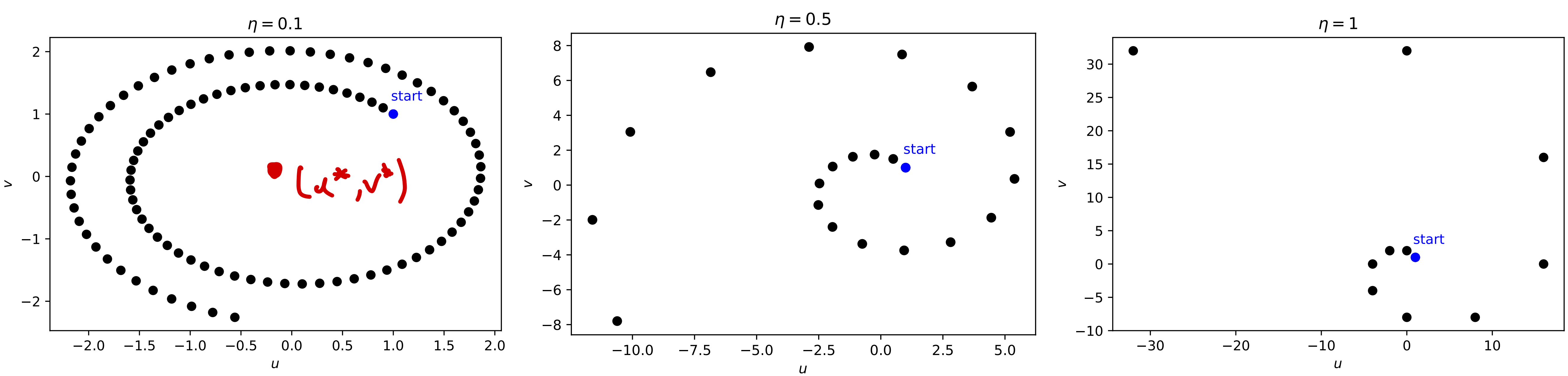
"Gradient" descent:

$$x_t = \arg\min_{x \in K} \left\{ \langle F(x_{t-1}), x \rangle + \frac{1}{2\eta} \| x - x_{t-1} \|^2 \right\}$$

Consider $\min_{u \in \mathbb{R}} \max_{v \in \mathbb{R}} uv$:

Equilibrium is $(u^*, v^*) = (0,0)$

start $= (u_0, v_0) = (1,1)$



# In Extra-Gradient Descent We Trust

Extra-"Gradient" Algorithm:

$$x_t = \arg\min_{u \in K} \left\{ \left\langle F(z_{t-1}), u \right\rangle + \frac{1}{2\eta} \| u - z_{t-1} \|^2 \right\}$$

$$z_t = \arg\min_{u \in K} \left\{ \left\langle F(x_t), u \right\rangle + \frac{1}{2\eta} \| u - z_{t-1} \|^2 \right\}$$

# In Extra-Gradient Descent We Trust

Extra-"Gradient" Algorithm:

$F = \nabla f$

$$x_t = \arg\min_{u \in K} \left\{ \langle F(z_{t-1}), u \rangle + \frac{1}{2\eta} \| u - z_{t-1} \|^2 \right\}$$

$$z_t = \arg\min_{u \in K} \left\{ \langle F(x_t), u \rangle + \frac{1}{2\eta} \| u - z_{t-1} \|^2 \right\}$$

Unconstrained ($K = \mathbb{R}^d$):

$$x_t = z_{t-1} - \eta F(z_{t-1})$$

$$z_t = z_{t-1} - \eta F(x_t) = z_{t-1} - \eta F\left(z_{t-1} - \eta F(z_{t-1})\right)$$

# Extra-Gradient Algorithm

## Extra-Gradient Algorithm (Korpelevich)

Let $z_0 \in K$

For $t = 1, \ldots, T$:

$$x_t = \arg\min_{u \in K} \left\{ \left\langle F(z_{t-1}), u \right\rangle + \frac{1}{2\eta} \| u - z_{t-1} \|^2 \right\}$$

$$z_t = \arg\min_{u \in K} \left\{ \left\langle F(x_t), u \right\rangle + \frac{1}{2\eta} \| u - z_{t-1} \|^2 \right\}$$

Return $\bar{x}_T = \frac{1}{T} \sum_{t=1}^{T} x_t$

# Extra-Gradient Analysis

We will analyze convergence via the error function:

$$\mathrm{Err}(x) := \sup_{y \in K} \langle F(y), x - y \rangle$$

Thus we want to upper bound $\mathrm{Err}(\bar{x}_T)$

Analogously to GD, we consider two settings:

"non-smooth" setting: $\|F(x)\| \leq G \quad \forall x \in K$

"smooth" setting: $\|F(x) - F(y)\| \leq \beta \|x - y\| \quad \forall x, y \in K$

# Extra-Gradient Analysis

We will analyze convergence via the error function:

$$\mathrm{Err}(x) := \sup_{y \in K} \langle F(y), x - y \rangle$$

Thus we want to upper bound $\mathrm{Err}(\bar{x}_T)$

Using that $F$ is monotone (analogue of convexity):

$$\mathrm{Err}(\bar{x}_T) = \sup_{y \in K} \left\langle F(y), \bar{x}_T - y \right\rangle \quad \text{definition}$$

$$= \sup_{y \in K} \left( \frac{1}{T} \sum_{t=1}^{T} \left\langle F(y), x_t - y \right\rangle \right) \quad \bar{x}_T = \frac{1}{T} \sum_{t=1}^{T} x_t$$

$$\leq \sup_{y \in K} \left( \frac{1}{T} \sum_{t=1}^{T} \left\langle F(x_t), x_t - y \right\rangle \right) \quad \text{monotonicity}$$

$$\langle F(x_t) - F(y), x_t - y \rangle \geq 0$$

# Extra-Gradient Analysis

We have: $\mathrm{Err}(\bar{x}_T) \leq \frac{1}{T} \sup_{y \in K} \left( \sum_{t=1}^{T} \langle F(x_t), x_t - y \rangle \right)$ Fix $y \in K$

We split each term so that we can use the optimality condition:

$$x_t = \arg\min_{u \in K} \left\{ \langle F(z_{t-1}), u \rangle + \frac{1}{2\eta} \| u - z_{t-1} \|^2 \right\}$$

$$z_t = \arg\min_{u \in K} \left\{ \langle F(x_t), u \rangle + \frac{1}{2\eta} \| u - z_{t-1} \|^2 \right\}$$

# Extra-Gradient Analysis

We have: $\mathrm{Err}(\bar{x}_T) \leq \frac{1}{T} \sup_{y \in K} \left( \sum_{t=1}^{T} \langle F(x_t), x_t - y \rangle \right)$

We split each term so that we can use the optimality condition:

$$\langle F(x_t), x_t - y \rangle = \langle F(x_t), z_t - y \rangle + \langle F(z_{t-1}), x_t - z_t \rangle + \langle F(x_t) - F(z_{t-1}), x_t - z_t \rangle$$

# Extra-Gradient Analysis

We have: $\mathrm{Err}(\bar{x}_T) \leq \frac{1}{T} \sup_{y \in K} \left( \sum_{t=1}^{T} \left\langle F(x_t), x_t - y \right\rangle \right)$

We split each term so that we can use the optimality condition:

$$\left\langle F(x_t), x_t - y \right\rangle = \left\langle F(x_t), z_t - y \right\rangle + \left\langle F(z_{t-1}), x_t - z_t \right\rangle + \left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle$$

We bound the first term using the optimality condition for $z_t$:

$$z_t = \arg\min_{u \in K} \left\{ \left\langle F(x_t), u \right\rangle + \frac{1}{2\eta} \| u - z_{t-1} \|^2 \right\}$$

# Extra-Gradient Analysis

We have: $\mathrm{Err}(\bar{x}_T) \leq \frac{1}{T} \sup_{y \in K} \left( \sum_{t=1}^{T} \langle F(x_t), x_t - y \rangle \right)$

We split each term so that we can use the optimality condition:

$$\langle F(x_t), x_t - y \rangle = \langle F(x_t), z_t - y \rangle + \langle F(z_{t-1}), x_t - z_t \rangle + \langle F(x_t) - F(z_{t-1}), x_t - z_t \rangle$$

We bound the first term using the optimality condition for $z_t$:

$$z_t = \arg\min_{u \in K} \left\{ \langle F(x_t), u \rangle + \frac{1}{2\eta} \| u - z_{t-1} \|^2 \right\}$$

$$\langle F(x_t), z_t - y \rangle \leq \frac{1}{\eta} \langle z_{t-1} - z_t, z_t - y \rangle$$

$ab = \frac{1}{2}(a+b)^2 - \frac{1}{2}a^2 - \frac{1}{2}b^2$

$$= \frac{1}{2\eta} \left( \| z_{t-1} - y \|^2 - \| z_t - y \|^2 - \| z_{t-1} - z_t \|^2 \right)$$

# Extra-Gradient Analysis

We have: $\mathrm{Err}(\bar{x}_T) \leq \frac{1}{T} \sup_{y \in K} \left( \sum_{t=1}^{T} \langle F(x_t), x_t - y \rangle \right)$

We split each term so that we can use the optimality condition:

$$\langle F(x_t), x_t - y \rangle = \langle F(x_t), z_t - y \rangle + \langle F(z_{t-1}), x_t - z_t \rangle + \langle F(x_t) - F(z_{t-1}), x_t - z_t \rangle$$

We bound the second term using the optimality condition for $x_t$:

$$x_t = \arg\min_{u \in K} \left\{ \langle F(z_{t-1}), u \rangle + \frac{1}{2\eta} \| u - z_{t-1} \|^2 \right\}$$

# Extra-Gradient Analysis

We have: $\mathrm{Err}(\bar{x}_T) \leq \frac{1}{T} \sup_{y \in K} \left( \sum_{t=1}^{T} \langle F(x_t), x_t - y \rangle \right)$

We split each term so that we can use the optimality condition:

$$\langle F(x_t), x_t - y \rangle = \langle F(x_t), z_t - y \rangle + \langle F(z_{t-1}), x_t - z_t \rangle + \langle F(x_t) - F(z_{t-1}), x_t - z_t \rangle$$

We bound the second term using the optimality condition for $x_t$:

$$x_t = \arg\min_{u \in K} \left\{ \langle F(z_{t-1}), u \rangle + \frac{1}{2\eta} \| u - z_{t-1} \|^2 \right\}$$

$$\langle F(z_{t-1}), x_t - z_t \rangle \leq \frac{1}{\eta} \langle z_{t-1} - x_t, x_t - z_t \rangle$$

$$= \frac{1}{2\eta} \left( \| z_{t-1} - z_t \|^2 - \| x_t - z_{t-1} \|^2 - \| x_t - z_t \|^2 \right)$$

# Extra-Gradient Analysis

We have:

$$\left\langle F(x_t), x_t - y \right\rangle = \left\langle F(x_t), z_t - y \right\rangle + \left\langle F(z_{t-1}), x_t - z_t \right\rangle + \left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle$$

$$\left\langle F(x_t), z_t - y \right\rangle \leq \frac{1}{2\eta} \left( \| z_{t-1} - y \|^2 - \| z_t - y \|^2 - \| z_{t-1} - z_t \|^2 \right)$$

+

$$\left\langle F(z_{t-1}), x_t - z_t \right\rangle \leq \frac{1}{2\eta} \left( \| z_{t-1} - z_t \|^2 - \| x_t - z_{t-1} \|^2 - \| x_t - z_t \|^2 \right)$$

# Extra-Gradient Analysis

We have:

$$\left\langle F(x_t), x_t - y \right\rangle = \left\langle F(x_t), z_t - y \right\rangle + \left\langle F(z_{t-1}), x_t - z_t \right\rangle + \left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle$$

$$\left\langle F(x_t), z_t - y \right\rangle \leq \frac{1}{2\eta}\left( \| z_{t-1} - y \|^2 - \| z_t - y \|^2 - \| z_{t-1} - z_t \|^2 \right)$$

$$\left\langle F(z_{t-1}), x_t - z_t \right\rangle \leq \frac{1}{2\eta}\left( \| z_{t-1} - z_t \|^2 - \| x_t - z_{t-1} \|^2 - \| x_t - z_t \|^2 \right)$$

Therefore

$$\left\langle F(x_t), x_t - y \right\rangle \leq \frac{1}{2\eta} \| z_{t-1} - y \|^2 - \frac{1}{2\eta} \| z_t - y \|^2$$

$$+ \left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle - \frac{1}{2\eta}\left( \| x_t - z_{t-1} \|^2 + \| x_t - z_t \|^2 \right)$$

# Extra-Gradient Analysis

We have:

$$\left\langle F(x_t), x_t - y \right\rangle = \left\langle F(x_t), z_t - y \right\rangle + \left\langle F(z_{t-1}), x_t - z_t \right\rangle + \left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle$$

$$\left\langle F(x_t), z_t - y \right\rangle \leq \frac{1}{2\eta}\left( \| z_{t-1} - y \|^2 - \| z_t - y \|^2 - \| z_{t-1} - z_t \|^2 \right)$$

$$\left\langle F(z_{t-1}), x_t - z_t \right\rangle \leq \frac{1}{2\eta}\left( \| z_{t-1} - z_t \|^2 - \| x_t - z_{t-1} \|^2 - \| x_t - z_t \|^2 \right)$$

Therefore

$$\left\langle F(x_t), x_t - y \right\rangle \leq \underbrace{\frac{1}{2\eta} \| z_{t-1} - y \|^2 - \frac{1}{2\eta} \| z_t - y \|^2}_{\text{telescopes}}$$

$$+ \underbrace{\left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle}_{\text{loss}} - \underbrace{\frac{1}{2\eta}\left( \| x_t - z_{t-1} \|^2 + \| x_t - z_t \|^2 \right)}_{\text{gain}}$$

# Extra-Gradient Analysis

Next, we analyze the net loss:

$$\underbrace{\left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle}_{\text{loss}} - \underbrace{\frac{1}{2\eta}\left( \| x_t - z_{t-1} \|^2 + \| x_t - z_t \|^2 \right)}_{\text{gain}}$$

# Extra-Gradient Analysis

Next, we analyze the net loss:

$$\underbrace{\left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle}_{\text{loss}} - \underbrace{\frac{1}{2\eta}\left( \| x_t - z_{t-1} \|^2 + \| x_t - z_t \|^2 \right)}_{\text{gain}}$$

In the "non-smooth" setting, we assume $\|F(x)\| \leq G$

We proceed similarly to the GD analysis, and obtain:

$$\left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle \leq \left\| F(x_t) - F(z_{t-1}) \right\| \| x_t - z_t \|$$ Cauchy-Schwartz

# Extra-Gradient Analysis

Next, we analyze the net loss:

$$\underbrace{\left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle}_{\text{loss}} - \underbrace{\frac{1}{2\eta}\left( \| x_t - z_{t-1} \|^2 + \| x_t - z_t \|^2 \right)}_{\text{gain}}$$

In the "non-smooth" setting, we assume $\|F(x)\| \leq G$

We proceed similarly to the GD analysis, and obtain:

$$\begin{aligned}\left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle &\leq \| F(x_t) - F(z_{t-1}) \| \, \| x_t - z_t \| \quad \text{Cauchy-Schwartz} \\ &\leq \left( \| F(x_t) \| + \| F(z_{t-1}) \| \right) \| x_t - z_t \| \quad \Delta\text{-ineq.}\end{aligned}$$

# Extra-Gradient Analysis

Next, we analyze the net loss:

$$\underbrace{\left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle}_{\text{loss}} - \underbrace{\frac{1}{2\eta}\left( \| x_t - z_{t-1} \|^2 + \| x_t - z_t \|^2 \right)}_{\text{gain}}$$

In the "non-smooth" setting, we assume $\|F(x)\| \leq G$

We proceed similarly to the GD analysis, and obtain:

$$\begin{aligned} \left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle &\leq \| F(x_t) - F(z_{t-1}) \| \, \| x_t - z_t \| \quad \text{Cauchy-Schwar} \\ &\leq \left( \| F(x_t) \| + \| F(z_{t-1}) \| \right) \| x_t - z_t \| \quad \Delta\text{-ineq.} \\ &\leq 2G \, \| x_t - z_t \| \end{aligned}$$

# Extra-Gradient Analysis

Next, we analyze the net loss:

$$\underbrace{\left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle}_{\text{loss}} - \underbrace{\frac{1}{2\eta}\left( \| x_t - z_{t-1} \|^2 + \| x_t - z_t \|^2 \right)}_{\text{gain}}$$

In the "non-smooth" setting, we assume $\|F(x)\| \leq G$

We proceed similarly to the GD analysis, and obtain:

$$\begin{aligned}
\left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle &\leq \| F(x_t) - F(z_{t-1}) \| \, \| x_t - z_t \| && \text{Cauchy-Schwartz} \\
&\leq \left( \| F(x_t) \| + \| F(z_{t-1}) \| \right) \| x_t - z_t \| && \Delta\text{-ineq.} \\
&\leq 2G \, \| x_t - z_t \| \\
&\leq 2\eta G^2 + \frac{1}{2\eta} \| x_t - z_t \|^2 && ab \leq \frac{\lambda}{2}a^2 + \frac{1}{2\lambda}b^2 \text{ for any } \lambda > 0
\end{aligned}$$

# Extra-Gradient Analysis

Putting everything together:

$$\langle F(x_t), x_t - y \rangle \leq \frac{1}{2\eta} \| z_{t-1} - y \|^2 - \frac{1}{2\eta} \| z_t - y \|^2 + 2\eta G^2$$

# Extra-Gradient Analysis

Putting everything together:

$$\left\langle F(x_t), x_t - y \right\rangle \leq \frac{1}{2\eta} \| z_{t-1} - y \|^2 - \frac{1}{2\eta} \| z_t - y \|^2 + 2\eta G^2$$

Summing up and telescoping:

$$\sum_{t=1}^{T} \left\langle F(x_t), x_t - y \right\rangle \leq \frac{1}{2\eta} \| z_0 - y \|^2 + 2\eta G^2 T$$

# Extra-Gradient Analysis

Putting everything together:

$$\langle F(x_t), x_t - y \rangle \leq \frac{1}{2\eta} \| z_{t-1} - y \|^2 - \frac{1}{2\eta} \| z_t - y \|^2 + 2\eta G^2$$

Summing up and telescoping:

$$\sum_{t=1}^{T} \langle F(x_t), x_t - y \rangle \leq \frac{1}{2\eta} \| z_0 - y \|^2 + 2\eta G^2 T$$

We set $\eta$ to balance the two terms:

$$\eta = \frac{\| z_0 - y \|}{2G\sqrt{T}}$$

# Extra-Gradient Analysis

Putting everything together:

$$\langle F(x_t), x_t - y \rangle \leq \frac{1}{2\eta} \| z_{t-1} - y \|^2 - \frac{1}{2\eta} \| z_t - y \|^2 + 2\eta G^2$$

Summing up and telescoping:

$$\sum_{t=1}^{T} \langle F(x_t), x_t - y \rangle \leq \frac{1}{2\eta} \| z_0 - y \|^2 + 2\eta G^2 T$$

We set $\eta$ to balance the two terms:

$$\eta = \frac{\| z_0 - y \|}{2G\sqrt{T}}$$

Thus

$$\sum_{t=1}^{T} \langle F(x_t), x_t - y \rangle \leq 2G \underbrace{\| z_0 - y \|}_{\leq R} \sqrt{T}$$

$$\leq 2GR\sqrt{T}$$

# Extra-Gradient Analysis

We have:

$$\mathrm{Err}(\bar{x}_T) \leq \frac{1}{T} \sup_{y \in K} \left( \sum_{t=1}^{T} \langle F(x_t), x_t - y \rangle \right)$$

$$\sum_{t=1}^{T} \langle F(x_t), x_t - y \rangle \leq 2GR\sqrt{T} \qquad \forall y \in K$$

Therefore we have our final convergence guarantee:

$$\mathrm{Err}(\bar{x}_T) \leq O\left( \frac{GR}{\sqrt{T}} \right)$$

# Extra-Gradient Analysis

Next, we consider the “smooth” (i.e., Lipschitz) setting:

$$\|F(x) - F(y)\| \leq \beta\|x - y\| \quad \forall x, y$$

# Extra-Gradient Analysis

Next, we consider the "smooth" (i.e., Lipschitz) setting:

$$\|F(x) - F(y)\| \leq \beta \|x - y\| \quad \forall x, y$$

As before, we need to analyze the net loss:

$$\underbrace{\left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle}_{\text{loss}} - \underbrace{\frac{1}{2\eta}\left( \| x_t - z_{t-1} \|^2 + \| x_t - z_t \|^2 \right)}_{\text{gain}}$$

# Extra-Gradient Analysis

Next, we consider the "smooth" (i.e., Lipschitz) setting:

$$\|F(x) - F(y)\| \leq \beta \|x - y\| \quad \forall x, y$$

As before, we need to analyze the net loss:

$$\underbrace{\left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle}_{\text{loss}} - \underbrace{\frac{1}{2\eta}\left( \| x_t - z_{t-1} \|^2 + \| x_t - z_t \|^2 \right)}_{\text{gain}}$$

Using the Lipschitz property, we obtain:

$$\left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle \leq \left\| F(x_t) - F(z_{t-1}) \right\| \; \| x_t - z_t \|$$ Cauchy-Schwartz

# Extra-Gradient Analysis

Next, we consider the "smooth" (i.e., Lipschitz) setting:

$$\|F(x) - F(y)\| \leq \beta \|x - y\| \quad \forall x, y$$

As before, we need to analyze the net loss:

$$\underbrace{\left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle}_{\text{loss}} - \underbrace{\frac{1}{2\eta}\left( \| x_t - z_{t-1} \|^2 + \| x_t - z_t \|^2 \right)}_{\text{gain}}$$

Using the Lipschitz property, we obtain:

$$\begin{aligned}\left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle &\leq \left\| F(x_t) - F(z_{t-1}) \right\| \; \| x_t - z_t \| && \text{CS} \\ &\leq \beta \; \| x_t - z_{t-1} \| \; \| x_t - z_t \| && \text{smoothness}\end{aligned}$$

# Extra-Gradient Analysis

Next, we consider the "smooth" (i.e., Lipschitz) setting:

$$\|F(x) - F(y)\| \leq \beta \|x - y\| \quad \forall x, y$$

As before, we need to analyze the net loss:

$$\underbrace{\langle F(x_t) - F(z_{t-1}), x_t - z_t \rangle}_{\text{loss}} - \underbrace{\frac{1}{2\eta}\left( \| x_t - z_{t-1} \|^2 + \| x_t - z_t \|^2 \right)}_{\text{gain}}$$

Using the Lipschitz property, we obtain:

$$\begin{aligned}
\langle F(x_t) - F(z_{t-1}), x_t - z_t \rangle &\leq \| F(x_t) - F(z_{t-1}) \| \, \| x_t - z_t \| && \text{C S} \\
&\leq \beta \, \| x_t - z_{t-1} \| \, \| x_t - z_t \| && \text{smoothness} \\
&\leq \frac{\beta}{2} \| x_t - z_{t-1} \|^2 + \frac{\beta}{2} \| x_t - z_t \|^2
\end{aligned}$$

# Extra-Gradient Analysis

As before, we need to analyze the net loss:

$$\underbrace{\langle F(x_t) - F(z_{t-1}), x_t - z_t \rangle}_{\text{loss}} - \underbrace{\frac{1}{2\eta}\left( \| x_t - z_{t-1} \|^2 + \| x_t - z_t \|^2 \right)}_{\text{gain}}$$

Using the Lipschitz property, we obtained:

$$\langle F(x_t) - F(z_{t-1}), x_t - z_t \rangle \leq \frac{\beta}{2} \| x_t - z_{t-1} \|^2 + \frac{\beta}{2} \| x_t - z_t \|^2$$

# Extra-Gradient Analysis

As before, we need to analyze the net loss:

$$\underbrace{\left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle}_{\text{loss}} - \underbrace{\frac{1}{2\eta}\left( \| x_t - z_{t-1} \|^2 + \| x_t - z_t \|^2 \right)}_{\text{gain}}$$

Using the Lipschitz property, we obtained:

$$\left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle \leq \frac{\beta}{2} \| x_t - z_{t-1} \|^2 + \frac{\beta}{2} \| x_t - z_t \|^2$$

Thus, if we set $\eta = \frac{1}{\beta}$, we obtain:

$$\underbrace{\left\langle F(x_t) - F(z_{t-1}), x_t - z_t \right\rangle}_{\text{loss}} - \underbrace{\frac{1}{2\eta}\left( \| x_t - z_{t-1} \|^2 + \| x_t - z_t \|^2 \right)}_{\text{gain}} \leq 0$$

# Extra-Gradient Analysis

Putting everything together:

$$\left\langle F(x_t), x_t - y \right\rangle \leq \frac{\beta}{2} \parallel z_{t-1} - y \parallel^2 - \frac{\beta}{2} \parallel z_t - y \parallel^2$$

# Extra-Gradient Analysis

Putting everything together:

$$\langle F(x_t), x_t - y \rangle \leq \frac{\beta}{2} \| z_{t-1} - y \|^2 - \frac{\beta}{2} \| z_t - y \|^2$$

Summing up and telescoping:

$$\sum_{t=1}^{T} \langle F(x_t), x_t - y \rangle \leq \frac{\beta}{2} \| z_0 - y \|^2 \leq \frac{\beta}{2} R^2$$

# Extra-Gradient Analysis

Putting everything together:

$$\left\langle F(x_t), x_t - y \right\rangle \leq \frac{\beta}{2} \parallel z_{t-1} - y \parallel^2 - \frac{\beta}{2} \parallel z_t - y \parallel^2$$

Summing up and telescoping:

$$\sum_{t=1}^{T} \left\langle F(x_t), x_t - y \right\rangle \leq \frac{\beta}{2} \parallel z_0 - y \parallel^2 \leq \frac{\beta}{2} R^2$$

Which gives us our final convergence guarantee:

$$\mathrm{Err}(\bar{x}_T) \leq O\left(\frac{\beta R^2}{T}\right)$$

# Extra-Gradient Analysis

Putting everything together:

$$\left\langle F(x_t), x_t - y \right\rangle \leq \frac{\beta}{2} \parallel z_{t-1} - y \parallel^2 - \frac{\beta}{2} \parallel z_t - y \parallel^2$$

Summing up and telescoping:

$$\sum_{t=1}^{T} \left\langle F(x_t), x_t - y \right\rangle \leq \frac{\beta}{2} \parallel z_0 - y \parallel^2 \leq \frac{\beta}{2} R^2$$

Which gives us our final convergence guarantee:

$$\mathrm{Err}(\bar{x}_T) \leq O\left(\frac{\beta R^2}{T}\right)$$

Both the “non-smooth” and “smooth” rates are optimal

# Extension to Bregman Divergences

## Mirror-Prox Algorithm (Nemirovski)

$\psi$ : strongly convex function

$D_\psi(x, y) = \psi(x) - \psi(y) - \langle \nabla \psi(y), x - y \rangle$

Let $z_0 \in K$

For $t = 1, \ldots, T$:

$$x_t = \arg\min_{u \in K} \left\{ \langle F(z_{t-1}), u \rangle + \frac{1}{2\eta} D_\psi(u, z_{t-1}) \right\}$$

$$z_t = \arg\min_{u \in K} \left\{ \langle F(x_t), u \rangle + \frac{1}{2\eta} D_\psi(u, z_{t-1}) \right\}$$

Return $\bar{x}_T = \frac{1}{T} \sum_{t=1}^{T} x_t$

# Adaptive Algorithms

## Adaptive Algorithm: Iterate Movement

Let $z_0 \in K, \eta_0 > 0, R \geq \max_{x,y \in K} \|x - y\|$

For $t = 1,\ldots,T$:

$$x_t = \arg\min_{u \in K} \left\{ \left\langle F(z_{t-1}), u \right\rangle + \frac{1}{2\eta_{t-1}} \| u - z_{t-1} \|^2 \right\}$$

$$z_t = \arg\min_{u \in K} \left\{ \left\langle F(x_t), u \right\rangle + \frac{1}{2\eta_{t-1}} \| u - z_{t-1} \|^2 \right\}$$

$$\frac{1}{\eta_t^2} = \frac{1}{\eta_{t-1}^2} \left( 1 + \frac{\| x_t - z_{t-1} \|^2 + \| x_t - z_t \|^2}{2R^2} \right)$$

Return $\bar{x}_T = \frac{1}{T} \sum_{t=1}^{T} x_t$

# Adaptive Algorithms

## Adaptive Algorithm: Operator Differences

Let $z_0 \in K, \eta_0 > 0, R \geq \max_{x,y \in K} \|x - y\|$

For $t = 1, \ldots, T$:

$$x_t = \arg\min_{u \in K} \left\{ \langle F(z_{t-1}), u \rangle + \frac{1}{2\eta_{t-1}} \| u - z_{t-1} \|^2 \right\}$$

$$\eta_t = \frac{R}{\sqrt{\sum_{s=1}^{t} \| F(x_s) - F(z_{s-1}) \|^2}}$$

$$z_t = \arg\min_{u \in K} \left\{ \langle F(x_t), u \rangle + \frac{1}{2\eta_{t-1}} \| u - z_{t-1} \|^2 \right\}$$

Return $\bar{x}_T = \frac{1}{T} \sum_{t=1}^{T} x_t$