Feedback welcome! If you find any typos or anything is unclear or misleading, please email me and know!

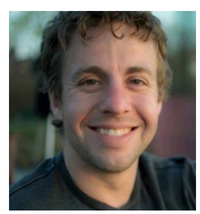
For additional detail, see the companion paper, “Box-Simplex Games : Algorithms, Applications, and Algorithmic Graph Theory” on my website.

Box-Simplex Games

Algorithms, Applications, and Algorithmic Graph Theory

Part 2

Aaron Sidford



Stanford University
Departments of Management Science
& Engineering and Computer Science

Contact Info:
- email: *sidford@stanford.edu*
- website: *www.aaronsidford.com*

The Problem (Recap)

Input

- n-dimensional box: $B_\infty^n \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid \|x\|_\infty \le 1\}$
- m-dimensional simplex: $\Delta^m \stackrel{\text{def}}{=} \{y \in \mathbb{R}_{\ge 0}^m \mid \|y\|_1 = 1\}$

Bounded vectors in $\mathbb{R}^n$

Probability distributions on m elements

Output:

- An approximate solution to

$$\min_{x \in B_\infty^n} \max_{y \in \Delta^m} f(x, y) \stackrel{\text{def}}{=} y^\top \boldsymbol{A} x + c^\top x - b^\top y$$

Box-Simplex Game $\quad$ ℓ_1-ℓ_∞ **Game**

Key Motivating Questions (Recap)

Question #1

How can we design efficient methods for solving box-simplex games?

Question #2

How can we leverage box-simplex solvers to solve continuous and combinatorial optimization problems?

Talk Plan (This Week)

- Box: $B_\infty^n \overset{\text{def}}{=} \{x \in \mathbb{R}^n \mid \|x\|_\infty \leq 1\}$
- Simplex: $\Delta^m \overset{\text{def}}{=} \{y \in \mathbb{R}^m_{\geq 0} \mid \|y\|_1 = 1\}$
- $\min_{x \in B_\infty^n} \max_{y \in \Delta^m} f(x, y) \overset{\text{def}}{=} y^\top \boldsymbol{A} x + c^\top x - b^\top y$

Part 1
Structure of box-simplex games

- Primal and dual problems
- Approximate solutions
- Discuss state-of-the-art runtimes

Friday
Interior Point Methods

As we have time.

Part 2
Applications

- Box-constrained ℓ_∞-regression
- Linear programming
- Maximum cardinality bipartite matching
- **Undirected maximum flow**

Part 3
Algorithms

- ℓ_∞-Gradient Descent (constrained steepest descent)
- ℓ_1-Mirror Descent (multiplicative weights)
- Mirror prox and primal dual regularizers

Approximate Solutions (Recap)

- Box: $B_\infty^n \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid \|x\|_\infty \le 1\}$
- Simplex: $\Delta^m \stackrel{\text{def}}{=} \{y \in \mathbb{R}^m_{\ge 0} \mid \|y\|_1 = 1\}$
- $\min_{x \in B_\infty^n} \max_{y \in \Delta^m} f(x,y) \stackrel{\text{def}}{=} y^\top \boldsymbol{A} x + c^\top x - b^\top y$

Primal Problem

- $\min_{x \in B_\infty^n} f_{\max}(x)$ and $f_{\max}(x) = \max_{y \in \Delta^m} f(x,y)$
- $f_{\max}(x) = c^\top x + \max_{i \in [m]} [\boldsymbol{A}x - b]_i$

Dual Problem

- $\max_{y \in \Delta^m} f_{\min}(y)$ and $f_{\min}(y) = \min_{x \in B_\infty^n} f(x,y)$
- $f_{\min}(y) = -b^\top y - \|\boldsymbol{A}^\top y - b\|_1$

Approximate Solutions

- ϵ-approximate *saddle-point* $(x_\epsilon, y_\epsilon) \in B_\infty^n \times \Delta^m$ if $f_{\max}(x_\epsilon) - f_{\min}(y_\epsilon) \le \epsilon$
  - $\Leftrightarrow \left[f_{\max}(x_\epsilon) - \min_{x \in B_\infty^n} f_{\max}(x)\right] + \left[\max_{y \in \Delta^m} f_{\min}(y) - f_{min}(y_\epsilon)\right] \le \epsilon$

Box-Simplex Solver Theorem

There is a method which solves box-simplex games
to accuracy ϵ in time $\tilde{O}\left(\mathrm{nnz}(\boldsymbol{A})\|\boldsymbol{A}\|_{\mathrm{op},\infty}/\epsilon\right)$.

Slightly different than last time and specialized for what we will prove.

Undirected Maximum Flow (Recap)

Theorem
There is an algorithm which can compute a $(1-\epsilon)$-approximate maximum flow, in time $\tilde{O}(|E|\epsilon^{-1})$

Capacitated Graph

- $G = (V, E, u)$
- Capacities: $u \in \mathbb{R}^E_{\geq 0}$

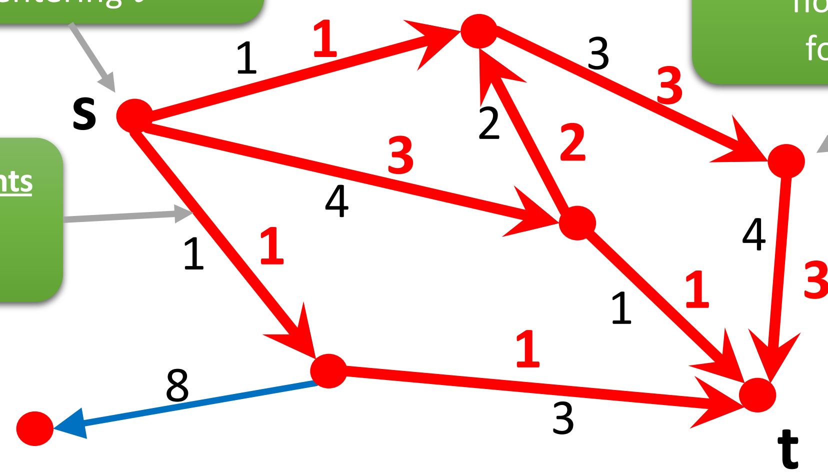
Value of Flow
total flow leaving s or entering t

$s \to t$ Flow
flow in = flow out for all $v \notin \{s, t\}$

Feasibility /Capacity Constraints
- Directed: $f_e \in [0, u_e]$
- Undirected: $f_e \in [-u_e, u_e]$

Terminals

- Source $s \in V$
- Sink $t \in V$



Goal
Compute $(1-\epsilon)$-approximate flow, i.e. a feasible $s \to t$ flow of value $\geq (1-\epsilon)OPT$.

Flow
$f \in \mathbb{R}^E$ where f_e = amount of flow on edge e

Linear Algebraic Formulation

Focus on undirected maximum flow but keep G directed. This is done so for $f \in \mathbb{R}^E$ meaning of f_e is clear.

- **Input**: Capacitated graph $G = (V, E, u)$ with $u \in \mathbb{R}^E_{\geq 0}$
- **Flow**: $f \in \mathbb{R}^E$
- **Imbalance**: $\mathrm{im}(f) \in \mathbb{R}^V$ with $[\mathrm{im}(f)]_a = \sum_{\{a,b\} \in E} f_{\{a,b\}} - \sum_{\{b,a\} \in E} f_{\{a,b\}}$
  - $\mathrm{im}(f) = \boldsymbol{B}^\top f$ for edge vertex incidence matrix $\boldsymbol{B} \in \mathbb{R}^{E \times V}$
- **$\boldsymbol{s}$-$\boldsymbol{t}$ flow of value $\boldsymbol{v}$**: $\mathrm{im}(f) = \nu \cdot \vec{\delta}_{s,t}$ with $\vec{\delta}_{s,t} \stackrel{\text{def}}{=} \vec{1}_s - \vec{1}_t$
- **$\boldsymbol{f} \in \mathbb{R}^{\boldsymbol{E}}$ satisfies capacity constraints**: $|f_e| \leq u_e$ for all $e \in E$
  - For $\boldsymbol{U} \stackrel{\text{def}}{=} \mathbf{diag}(u)$ this is the same as $\|\boldsymbol{U}^{-1} f\|_\infty \leq 1$
- **Linear algebraic formulation**: $$\max_{f \in \mathbb{R}^E | \boldsymbol{B}^\top f = \alpha \vec{\delta}_{s,t} \text{ and } \|\boldsymbol{U}^{-1} f\|_\infty \leq 1} \alpha$$

$$\boldsymbol{B}_{(a,b),c} = \begin{cases} 1 & a = c \\ -1 & b = c \\ 0 & otherwise \end{cases}$$ for all $(a, b) \in E$ and $c \in V$

Equivalent Problem

Demands: $d \in \mathbb{R}^V$ where $\exists f \in \mathbb{R}^E$ with $\boldsymbol{B}^\top f = d$.
For connected undirected graph equivalent to $d \perp 1$

Congestion: $\|\boldsymbol{U}^{-1} f\|_\infty$

Maximum Flow

- **Input**: $G = (V, E, u)$ and $s, t \in V$
- **Problem**: $\alpha_* = \max_{f \in \mathbb{R}^E \mid \boldsymbol{B}^\top f = \alpha \vec{\delta}_{s,t} \text{ and } \|\boldsymbol{U}^{-1} f\|_\infty \leq 1} \alpha$
- **$(1-\epsilon)$-approximate**: $f_\epsilon \in \mathbb{R}^E$ with $\boldsymbol{B}^\top f_\epsilon = \alpha \vec{\delta}_{s,t}$, $\|\boldsymbol{U}^{-1} f_\epsilon\|_\infty \leq 1$, and $\alpha \geq (1-\epsilon)\alpha_*$

Minimum Congestion Flow

- **Input**: $G = (V, E, u)$ and demands $d \in \mathbb{R}^V$
- **Problem**: $\mathrm{OPT}(d) \stackrel{\text{def}}{=} \min_{f \in \mathbb{R}^E \mid \boldsymbol{B}^\top f = d} \|\boldsymbol{U}^{-1} f\|_\infty$
- **$(1+\epsilon)$-approximate**: $f_\epsilon \in \mathbb{R}^E$ with $\boldsymbol{B}^\top f = d$ and $\|\boldsymbol{U}^{-1} f\|_\infty \leq (1+\epsilon)\mathrm{OPT}(d)$

Lemma: $\mathrm{OPT}(\vec{\delta}_{s,t}) = \frac{1}{\alpha_*}$ and given any $(1+\epsilon)$-approximate minimum congestion flow f_ϵ for demands $\vec{\delta}_{s,t}$ it is the case that $\frac{1}{\|\boldsymbol{U}^{-1} f\|_\infty} f_\epsilon$ is a $\frac{1}{1+\epsilon}$-approximate maximum flow.

Proof:

- Let f_* be a maximum flow. Then $\mathbf{B}^\top \left(\frac{1}{\alpha_*} f_*\right) = \vec{\delta}_{s,t}$ with $\left\| U^{-1} \left(\frac{1}{\alpha_*} f_*\right)\right\|_\infty \leq \frac{1}{\alpha_*}$ and therefore $\mathrm{OPT}(\vec{\delta}_{s,t}) \leq \frac{1}{\alpha_*}$
- Note that $\boldsymbol{B}^\top \left(\frac{1}{\|\boldsymbol{U}^{-1} f_\epsilon\|_\infty} f_\epsilon\right) = \frac{1}{\|\boldsymbol{U}^{-1} f_\epsilon\|_\infty} \geq \frac{1}{(1+\epsilon)\mathrm{OPT}(\vec{\delta}_{s,t})} \geq \frac{\alpha_*}{1+\epsilon}$

Since $\frac{1}{1+\epsilon} \approx 1 - \epsilon$ if suffices to solve minimum congestion flow!

Almost Equivalent Problem

Minimum Congestion Flow

- **Input**: $G = (V, E, u)$ and demands $d \in \mathbb{R}^V$
- **Problem**: $\mathrm{OPT}(d) \stackrel{\text{def}}{=} \min_{f \in \mathbb{R}^E | \boldsymbol{B}^\top f = d} \|\boldsymbol{U}^{-1} f\|_\infty$
- $(\mathbf{1} + \boldsymbol{\epsilon})$**-approximate**: $f_\epsilon \in \mathbb{R}^E$ with $\boldsymbol{B}^\top f = d$ and $\|\boldsymbol{U}^{-1} f\|_\infty \leq (1 + \epsilon)\mathrm{OPT}(d)$

Flow Feasibility

- **Input**: $G = (V, E, u)$ and demands $d \in \mathbb{R}^V$
- **Promise**: $\exists f \in \mathbb{R}^E$ with $\boldsymbol{B}^\top f = d$ and $\|\boldsymbol{U}^{-1} f\|_\infty \leq 1$
- **Output**: $f \in \mathbb{R}^E$ with $\boldsymbol{B}^\top f = d$ and $\|\boldsymbol{U}^{-1} f\|_\infty \leq 1 + \epsilon$

Reduction

- Solving flow feasibility for $d \coloneqq \frac{1}{\mathrm{OPT}(d)} d$ suffices
- Can find a good enough scaling of d by binary search
- Result: suffices to solve $O(\log \epsilon^{-1})$ flow feasibility problems for $\epsilon \coloneqq \frac{\epsilon}{2}$

Approach?

Minimum Congestion Flow

- **Input**: $G = (V, E, u)$ and demands $d \in \mathbb{R}^V$
- **Problem**: $\mathrm{OPT}(d) \stackrel{\text{def}}{=} \min_{f \in \mathbb{R}^E | \boldsymbol{B}^\top f = d} \|\boldsymbol{U}^{-1} f\|_\infty$
- **$(\boldsymbol{1} + \boldsymbol{\epsilon})$-approximate**: $f_\epsilon \in \mathbb{R}^E$ with $\boldsymbol{B}^\top f = d$ and $\|\boldsymbol{U}^{-1} f\|_\infty \leq (1 + \epsilon)\mathrm{OPT}(d)$

Flow Feasibility

- **Input**: $G = (V, E, u)$ and demands $d \in \mathbb{R}^V$
- **Promise**: $\exists f \in \mathbb{R}^E$ with $\boldsymbol{B}^\top f = d$ and $\|\boldsymbol{U}^{-1} f\|_\infty \leq 1$
- **Output**: $f \in \mathbb{R}^E$ with $\boldsymbol{B}^\top f = d$ and $\|\boldsymbol{U}^{-1} f\|_\infty \leq 1 + \epsilon$

Idea

- Pick penalty function p and solve $\min_{f \in \mathbb{R}^E \mid \|\boldsymbol{U}^{-1} f\|_\infty \leq 1} p(\boldsymbol{B}^\top f - d)$
- Same as $\min_{x \in B_\infty^E} p(\boldsymbol{B}^\top \boldsymbol{U} x - d)$
- What p to pick?
- Box-constrained ℓ_∞ regression? $p(z) = \|z\|_\infty$?
  - Problems! How bound $\|\boldsymbol{B}^\top \boldsymbol{U}\|_\infty$? How to handle approximation error?

Congestion Approximators

- Find a penalty function better capturing problem
- Part of broader theory involving preconditioning

Congestion Approximator: $\boldsymbol{R} \in \mathbb{R}^{k \times V}$ is an α-congestion approximator if $\|\boldsymbol{R}d\|_\infty \leq OPT(d) \leq \alpha \cdot \|\boldsymbol{R}d\|_\infty$ for all demands $d \in \mathbb{R}^V$.

Theorem [P16]: there is an algorithm which given any capacitated undirected $G = (V, E, u)$ in $\tilde{O}(|E|)$ computes an $\tilde{O}(1)$-congestion approximator with $k = O(|V|)$, each column of $\boldsymbol{R}$ being $\tilde{O}(1)$ sparse and $\|\alpha \boldsymbol{R}\boldsymbol{B}^\top \boldsymbol{U}\|_\infty = \tilde{O}(1)$.

Approach

Recurse until residual demand is so small can solve naively!

Can choose parameters so bottleneck is the first solve up to an additive term.

Yields $\tilde{O}(|E|\epsilon^{-1})$ maximum flow!

Congestion Approximator: $\boldsymbol{R} \in \mathbb{R}^{k \times V}$ is an α-congestion approximator if $\|\boldsymbol{R}d\|_\infty \leq OPT(d) \leq \alpha \cdot \|\boldsymbol{R}d\|_\infty$ for all demands $d \in \mathbb{R}^V$.

Theorem [P16]: there is an algorithm which given any capacitated undirected $G = (V, E, u)$ in $\tilde{O}(|E|)$ computes an $\tilde{O}(1)$-congestion approximator with $k = O(|V|)$, each column of R being $\tilde{O}(1)$ sparse and $\|\alpha \boldsymbol{R} B^\top \boldsymbol{U}\|_\infty = 1$ in time $\tilde{O}(|E|)$.

Approach [S13,S17]: solve

$$\min_{\|\boldsymbol{U}^{-1} f\|_\infty \leq 1} \|\alpha \boldsymbol{R}(\boldsymbol{B}^T f - d)\|_\infty = \min_{\|x\|_\infty \leq 1} \|\alpha \boldsymbol{R}(\boldsymbol{B}^T \boldsymbol{U} x - d)\|_\infty$$

- Value of minimizer is 0 and consequently, can obtain ϵ function error in $\tilde{O}(|E|\epsilon^{-1})$ time using box-constrained regression algorithm!
- $\Rightarrow$ obtain f with $\|\boldsymbol{U}^{-1} f\|_\infty \leq 1$ and $OPT(\boldsymbol{B}^\top f - d) \leq \epsilon$ in time $\tilde{O}(|E|\epsilon^{-1})$
- If find any f' with $\|\boldsymbol{U}^{-1} f'\|_\infty \leq 2\epsilon$ and $\boldsymbol{B}^\top f' = \boldsymbol{B}^\top f - d$ then obtain $1 + 2\epsilon$ approximation.

Talk Plan (This Week)

- Box: $B_\infty^n \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid \|x\|_\infty \leq 1\}$
- Simplex: $\Delta^m \stackrel{\text{def}}{=} \{y \in \mathbb{R}^m_{\geq 0} \mid \|y\|_1 = 1\}$
- $\min_{x \in B_\infty^n} \max_{y \in \Delta^m} f(x, y) \stackrel{\text{def}}{=} y^\top \boldsymbol{A} x + c^\top x - b^\top y$

Part 1
Structure of box-simplex games

- Primal and dual problems
- Approximate solutions
- Discuss state-of-the-art runtimes

Friday
Interior Point Methods

As we have time.

Part 2
Applications

- Box-constrained ℓ_∞-regression
- Linear programming
- Maximum cardinality bipartite matching
- Undirected maximum flow

Part 3
Algorithms

- ℓ_∞-Gradient Descent (constrained steepest descent)
- ℓ_1-Mirror Descent (multiplicative weights)
- Mirror prox and primal dual regularizers

Approximate Solutions (Recap)

- Box: $B_\infty^n \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid \|x\|_\infty \le 1\}$
- Simplex: $\Delta^m \stackrel{\text{def}}{=} \{y \in \mathbb{R}_{\ge 0}^m \mid \|y\|_1 = 1\}$
- $\min_{x \in B_\infty^n} \max_{y \in \Delta^m} f(x,y) \stackrel{\text{def}}{=} y^\top \boldsymbol{A} x + c^\top x - b^\top y$

Primal Problem

- $\min_{x \in B_\infty^n} f_{\max}(x) = \max_{y \in \Delta^m} f(x,y)$
- $f_{\max}(x) = c^\top x + \max_{i \in [m]} [\boldsymbol{A}x - b]_i$

Dual Problem

- $\max_{x \in B_\infty^n} f_{\min}(y) = \min_{y \in \Delta^m} f(x,y)$
- $f_{\min}(y) = -b^\top y - \|\boldsymbol{A}^\top y - b\|_1$

Approximate Solutions

- ϵ-approximate *saddle-point* $(x_\epsilon, y_\epsilon) \in B_\infty^n \times \Delta^m$ if $f_{\max}(x_\epsilon) - f_{\min}(y_\epsilon) \le \epsilon$
- $\Leftrightarrow \left[f_{\max}(x_\epsilon) - \min_{x \in B_\infty^n} f_{\max}(x)\right] + \left[\max_{y \in \Delta^m} f_{\min}(y) - f_{min}(y_\epsilon)\right] \le \epsilon$

Box-Simplex Solver Theorem

There is a method which solves box-simplex games to accuracy ϵ in time $\tilde{O}\left(\mathrm{nnz}(\boldsymbol{A}) \|\boldsymbol{A}\|_{\mathrm{op},\infty} \epsilon^{-1}\right)$.

Warmup Algorithms

- Box: $B_\infty^n \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid \|x\|_\infty \leq 1\}$
- Simplex: $\Delta^m \stackrel{\text{def}}{=} \{x \in \mathbb{R}_{\geq 0}^m \mid \|x\|_1 = 1\}$
- $\min_{x \in B_\infty^n} \max_{y \in \Delta^m} f(x,y) \stackrel{\text{def}}{=} y^\top \boldsymbol{A} x + c^\top x - b^\top y$

Primal Problem

- $\min_{x \in B_\infty^n} f_{\max}(x) = \max_{y \in \Delta^m} f(x,y)$
- $f_{\max}(x) = c^\top x + \max_{i \in [m]} [\boldsymbol{A}x - b]_i$

Dual Problem

- $\max_{x \in B_\infty^n} f_{\min}(y) = \min_{y \in \Delta^m} f(x,y)$
- $f_{\min}(y) = -b^\top y - \|\boldsymbol{A}^\top y - b\|_1$

ℓ_∞-Gradient Descent

- $\sim\epsilon^{-2}$ iteration method

Mirror Descent in ℓ_1

- $\sim\epsilon^{-2}$ iteration method

Norm Duality

For any norm $\|\cdot\|: \mathbb{R}^n \to \mathbb{R}$ the dual $\|\cdot\|_*$ is defined for all $x \in \mathbb{R}^n$ as $\|x\|_* = \max_{\|y\| \leq 1} y^\top x$.

Dual Norm Facts

- $\|\cdot\|_1$ is the dual norm for $\|\cdot\|_\infty$ and $\|\cdot\|_\infty$ is the dual norm for $\|\cdot\|_1$.
- For any norm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ its dual norm $\|\cdot\|_*$ is a norm
- $|x^\top y| \leq \|x\| \cdot \|y\|_*$ for al $x, y \in \mathbb{R}^n$

Gradient Descent

Primal Problem

- Box: $B_\infty^n \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid \|x\|_\infty \le 1\}$
- $\min_{x \in B_\infty^n} f_{\max}(x) = c^\top x + \max_{i \in [m]} [\boldsymbol{A}x - b]_i$

Basic Method

- For differentiable $f: \mathbb{R}^n \to \mathbb{R}$ iterate
$$x_{k+1} = x_k - \eta_k \nabla f(x_k)$$

Smoothness

- f is L-smooth if for all $x, y \in \mathbb{R}^n$
$$\|\nabla f(x) - \nabla f(y)\|_2 \le L \cdot \|x - y\|_2$$

Lemma: for convex, L-smooth f, and $\eta = \frac{1}{L}$
$$f(x_k) - \inf_x f(x) = O\left(\frac{L\|x_0 - x_*\|_2^2}{k}\right)$$

Geometric Motivation

- For L-smooth $f: \mathbb{R}^n \to \mathbb{R}$ and all $x, y \in \mathbb{R}^n$
$$f(y) \le U_x(y) \stackrel{\text{def}}{=} f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2}\|y - x\|_2^2$$

- Gradient descent derivation
$$x_k - \eta_k \nabla f(x_k) = \operatorname{argmin}_{x \in \mathbb{R}^n} U_{x_k}(x)$$

Obstacle Towards Applying to Primal Problem

- Problem is non-smooth (its non-differentiable)
- Problem is constrained
- $\|x_0 - x_*\|_2^2$ can be $\Omega(n)$

How to handle that problem is non-smooth

Overcoming Obstacles

Primal Problem

- Box: $B_\infty^n \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid \|x\|_\infty \leq 1\}$
- $\min_{x \in B_\infty^n} f_{\max}(x) = c^\top x + \max_{i \in [m]} [\boldsymbol{A}x - b]_i$

Problem is non-smooth

- Smooth it!
- $\text{smax}_t(x) \stackrel{\text{def}}{=} t \cdot \ln(\sum_i \exp(x_i/t))$
- $\tilde{f}_{\max,t}(x) = c^\top x + \text{smax}_t(\boldsymbol{A}x - b)$

Problem is constrained

- Incorporate the constraints!
- $x_{k+1} = \text{argmin}_{x \in B_\infty^n} U_{x_k}(x)$
- $U_{x_k}(y) \stackrel{\text{def}}{=} f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2} \|y - x\|_2^2$

Lem: twice differentiable $f: \chi \to \mathbb{R}$ is convex and L-smooth with respect to $\|\cdot\| \Leftrightarrow$ for all $x, z \in \chi$

Helpful for analyzing f. $0 \leq z^\top \nabla f(x) z \leq \frac{L}{2} \|z\|^2$

ℓ_2 versus ℓ_∞ ($\|x_0 - x_*\|_2^2$ can be $\Omega(n)$)

- Work in ℓ_∞!
- $x_{k+1} = \text{argmin}_{x \in B_\infty^n} U_{x_k}^\infty(x)$
- $U_{x_k}(y) \stackrel{\text{def}}{=} f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2} \|y - x\|_\infty^2$

Def: $f: \chi \to \mathbb{R}$ is L-smooth with respect to $\|\cdot\| \Leftrightarrow$

$\|\nabla f(x) - \nabla f(y)\|_* \leq L \cdot \|x - y\|$ for all $x, y \in \chi$

Helpful for analyzing algorithm.

Lem: $f: \chi \to \mathbb{R}$ is convex and L-smooth with respect to $\|\cdot\| \Leftrightarrow$ for all $x, y \in \chi$

$$0 \leq f(y) - [f(x) + \nabla f(x)^\top (y - x)] \leq \frac{L}{2} \|y - x\|_2^2$$

Smoothed Primal

Primal Problem

- Box: $B_\infty^n \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid \|x\|_\infty \le 1\}$
- $\min_{x \in B_\infty^n} f_{\max}(x) = c^\top x + \max_{i\in[m]} [\boldsymbol{A}x - b]_i$
- $\text{smax}_t(x) \stackrel{\text{def}}{=} t \cdot \ln(\sum_i \exp(x_i/t))$
- $\tilde{f}_{\max,t}(x) = c^\top x + \text{smax}_t(\boldsymbol{A}x - b)$

Lemma: $\text{smax}_t \colon \mathbb{R}^n \to \mathbb{R}$ for all $x \in \mathbb{R}^n$ is

- convex
- t^{-1} smooth with respect to $\|\cdot\|_\infty$
- $0 \le \text{smax}_t(x) - \max_{i\in[n]} x_i \le t \ln n$
- $[\nabla \text{smax}_t(x)]_i = \frac{\exp(x_i/t)}{\sum_{i\in[n]} \exp(x_i/t)}$

Corollary: $\tilde{f}_{\max,t}$ for all $x \in B_\infty^n$ is

- Convex
- $\|\boldsymbol{A}\|_\infty^2 t^{-1}$-smooth
- $0 \le \tilde{f}_{\max,t}(x) - f_{\max}(x) \le t \ln n$
- Gradient of $\tilde{f}_{\max,t}$ computable in $O(\text{nnz}(\boldsymbol{A}))$

Proof let $e_i = \exp(x_i/t)$ and $g_i = e_i/\|e\|_1$

- $\max_{i\in[n]} e_i \le \sum_{i\in[n]} e_i \le n \cdot \max_{i\in[n]} e_i$
- $\frac{\partial}{\partial x_i} \text{smax}_t(x) = \frac{e_i}{\|e\|_1}$
- $\frac{\partial^2}{\partial x_i \partial x_j} \text{smax}_t(x) = \frac{1}{t}\left[\frac{1_{i=j} e_i}{\|e\|_1} - \frac{e_i e_j}{\|e\|_1^2}\right]$
- $\nabla \text{smax}_t(x) = g$ and $\nabla^2 \text{smax}_t(x) = \frac{1}{t}[G - gg^\top]$ where $\boldsymbol{G} = \textbf{diag}(g)$.
- $(g^\top z)^2 = \left[\sum_{i\in[n]} \sqrt{g_i} \sum_{i\in[n]} \sqrt{g_i}\, z_i\right]^2$
  $\le \left(\sum_{i\in[n]} g_i\right)\left[\sum_{i\in[n]} g_i z_i^2\right] = z^\top \boldsymbol{G} z$
- $z^\top \boldsymbol{G} z = \sum_{i\in[n]} g_i z_i^2 \le \|g\|_1 \|z\|_\infty^2$
- $\nabla \tilde{f}_{\max,t}(x) = c + \boldsymbol{A}^\top \nabla \text{smax}_t(\boldsymbol{A}x - b)$
- $\nabla^2 \tilde{f}_{\max,t}(x) = \boldsymbol{A}^\top \nabla^2 \text{smax}_t(\boldsymbol{A}x - b)\boldsymbol{A}$

Constrained ℓ_∞-Gradient Descent

Proximal Point Method

$$x_{k+1} \in \underset{x \in \chi}{\operatorname{argmin}}\, f(x) + \frac{L}{2}\|x - x_k\|^2$$

Setup

- $f: \chi \to \mathbb{R}$ for closed, convex $\chi \subseteq \mathbb{R}^n$
- f is convex and
- f is L-smooth with respect to $\|\cdot\|$

Method

- $U_{x_k}(x) \stackrel{\text{def}}{=} f(x_k) + \nabla f(x_k)^\top (x - x_k) + \frac{L}{2}\|x - x_k\|^2$
- $x_{k+1} = \operatorname{argmin}_{x \in \chi} U_{x_k}(x)$for arbitrary $x_0 \in \chi$

Analysis

- $f(x_{k+1}) \leq U_{x_k}(x_{k+1})$ [smoothness]
- $U_{x_k}(x_{k+1}) = \min_{x \in \chi} U_{x_k}(x)$ [algorithm]
- $f(x) \geq f(x_k) + \nabla f(x_k)^\top (x - x_k)$ [convexity]
- $f(x_{k+1}) \leq \min_{x \in \chi} f(x) + \frac{L}{2}\|x - x_k\|^2$

Proximal Point Progres Lemma

- closed, convex χ, convex $f: \chi \to \mathbb{R}$, and norm $\|\cdot\|$
- $g(x) \stackrel{\text{def}}{=} f(x) + \frac{L}{2}\|x - x_0\|^2$ for all $x \in \chi$ and some $x_0 \in \chi$
- $x_* \in \operatorname{argmin}_{x \in \chi} f(x)$, $\Delta \stackrel{\text{def}}{=} f(x_0) - f(x_*)$, $D = \|x_0 - x_*\|$

$\Rightarrow \min_{x \in \chi} g(x) \leq f(x_0) - \frac{\Delta}{2} \cdot \min\left\{1, \frac{\Delta}{LD^2}\right\}$

Proof: Let $x_t = (1 - t) \cdot x_0 + t \cdot x_*$ for all $t \in [0,1]$

- $\|x_t - x_0\| = \|t \cdot (x_* - x_0)\| = tD$
- $f(x_t) \leq (1 - t) \cdot f(x_0) + t \cdot f(x_*) = f(x_0) - t\Delta$
- $g(x_t) = f(x_t) + \frac{L}{2}\|x_t - x_0\|^2 \leq f(x_0) - t\Delta + \frac{L}{2} t^2 D^2$
- If $t_* = \frac{\Delta}{LD^2} \in [0,1]$ then $g\left(x_{t_*}\right) \leq f(x_0) - \frac{\Delta^2}{2LD^2}$
- Otherwise $t_* > 1$, $\Delta > LD^2$, and $g(x_1) \leq f(x_0) - \frac{\Delta}{2}$

Constrained ℓ_∞-Gradient Descent

Setup

- $f: \chi \to \mathbb{R}$ for closed, convex $\chi \subseteq \mathbb{R}^n$
- f is convex and
- f is L-smooth with respect to $\|\cdot\|$

Method

- $U_{x_k}(x) \stackrel{\text{def}}{=} f(x_k) + \nabla f(x_k)^\top (x - x_k) + \frac{L}{2}\|x - x_k\|^2$
- $x_{k+1} = \text{argmin}_{x\in\chi} U_{x_k}(x)$for arbitrary $x_0 \in \chi$

Analysis

- $f(x_{k+1}) \leq \min_{x\in\chi} f(x) + \frac{L}{2}\|x - x_k\|_\infty^2$
- $\Delta_k = f(x_k) - \inf_{x\in\chi} f(x)$, $D = \max_{x,y\in\chi} \|x - y\|$
- $f(x_{k+1}) \leq f(x_k) - \frac{\Delta_k}{2} \cdot \min\left\{1, \frac{\Delta_k}{LD^2}\right\}$

Theorem

- $\Delta_k \leq \frac{LD^2}{k+1}$ for all $k \geq 1$

Proof:

- Δ_k decreases monotonically
- $f(x_{k+1}) \leq f(x_*) + \frac{L}{2}\|x_* - x_k\|_\infty^2$ so $\Delta_k \leq \frac{L}{2}D^2$ for all $k \geq 1$
- For all $k \geq 1$

$$\frac{1}{\Delta_k} - \frac{1}{\Delta_{k+1}} = \frac{\Delta_{k+1} - \Delta_k}{\Delta_k \Delta_{k+1}} \leq -\frac{1}{2\Delta_{k+1}} \min\left\{1, \frac{\Delta_k}{LD^2}\right\} \leq \frac{-1}{2LD^2}$$

- Summing and using that $\Delta_1 \leq \frac{L}{2}D^2$ yields

$$\frac{1}{\Delta_1} - \frac{1}{\Delta_{k+1}} \leq -\frac{k}{2LD^2} \text{ and } \frac{1}{\Delta_{k+1}} \geq \frac{k+4}{2LD^2}$$

Note: depends on the norm. improvability

A Primal ϵ^{-2} algorithm

Primal Problem

- Box: $B_\infty^n \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid \|x\|_\infty \leq 1\}$
- $\min_{x \in B_\infty^n} f_{\max}(x) = c^\top x + \max_{i \in [m]} [\boldsymbol{A}x - b]_i$
- $\text{smax}_t(x) \stackrel{\text{def}}{=} t \cdot \ln(\sum_i \exp(x_i/t))$
- $\tilde{f}_{\max,t}(x) = c^\top x + \text{smax}_t(\boldsymbol{A}x - b)$

Recall (Structure)

- $\tilde{f}_{\max,t}$ is convex and $\|\boldsymbol{A}\|_\infty^2 t^{-1}$-smooth
- $0 \leq \tilde{f}_{\max,t}(x) - f_{\max}(x) \leq t \ln n$
- Gradient of $\tilde{f}_{\max,t}$ computable in $O(\text{nnz}(\boldsymbol{A}))$

Recall (Algorithm)

- $U_{x_k}(x) \stackrel{\text{def}}{=} f(x_k) + \nabla f(x_k)^\top (x - x_k) + \frac{L}{2}\|x - x_k\|^2$
- $x_{k+1} = \text{argmin}_{x \in \chi} U_{x_k}(x)$ for $x_0 \in \chi$
- $\Delta_k = f(x_k) - \inf_{x \in \chi} f(x)$, $D = \max_{x,y \in \chi} \|x - y\|$
- $\Delta_k \leq \frac{LD^2}{k+1}$ for all $k \geq 1$

Analysis $\tilde{O}(\text{nnz}(\boldsymbol{A})\|\boldsymbol{A}\|_{op,\infty}^2 \epsilon^{-2})$

- Pick $t = \frac{\epsilon}{2 \log n}$
- $\tilde{f}_{\max,t}(x_k) - \min_{x \in B_\infty^n} \tilde{f}_{\max,t}(x) \leq \frac{4 \log n \|\boldsymbol{A}\|_\infty^2}{\epsilon k}$
- Pick $k = \lceil \frac{8 \log n \|\boldsymbol{A}\|_\infty^2}{\epsilon^2} \rceil$
- $f_{\max}(x_k) \leq f_{\max}(x_*) + t \log n + \frac{\epsilon}{2} \leq f_{\max}(x_*) + \epsilon$
- Algorithm implementation?
  - Can implement each step in $\tilde{O}(\text{nnz}(\boldsymbol{A}))$

Warmup Algorithms

- Box: $B_\infty^n \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid \|x\|_\infty \leq 1\}$
- Simplex: $\Delta^m \stackrel{\text{def}}{=} \{x \in \mathbb{R}^m_{\geq 0} \mid \|x\|_1 = 1\}$
- $\min_{x \in B_\infty^n} \max_{y \in \Delta^m} f(x,y) \stackrel{\text{def}}{=} y^\top \boldsymbol{A} x + c^\top x - b^\top y$

Primal Problem

- $\min_{x \in B_\infty^n} f_{\max}(x)$
- $f_{\max}(x) = \max_{y \in \Delta^m} f(x,y)$
- $f_{\max}(x) = c^\top x + \max_{i \in [m]} [\boldsymbol{A}x - b]_i$

Dual Problem

- $\max_{y \in \Delta^m} f_{\min}(y)$
- $f_{\min}(y) = \min_{x \in B_\infty^n} f(x,y)$
- $f_{\min}(y) = -b^\top y - \|\boldsymbol{A}^\top y - b\|_1$

ℓ_∞-Gradient Descent

- $\sim\epsilon^{-2}$ iteration method

Mirror Descent in ℓ_1

- $\sim\epsilon^{-2}$ iteration method

Subgradient Descent

Dual Problem

- Simplex: $\Delta^m \stackrel{\text{def}}{=} \{x \in \mathbb{R}^m_{\geq 0} \mid \|x\|_1 = 1\}$
- $\min_{y \in \Delta^m} -f_{\min}(y) = b^\top y + \|\boldsymbol{A}^\top y - b\|_1$

Part of broader theory

Subgradient: g is a subgradient of $f: \chi \to \mathbb{R}$ at $x \in \chi$ if $f(y) \geq f(x) + g^\top(y - x)$.

Subgradient Descent

- $x_{t+1} = x_t - \eta_t g_t$ for $g_t \in \partial f_{\min}(x_t)$
- Output $\bar{x}_T = \frac{1}{T}\sum_{t \in [T]} x_t$
- $\|g_t\|_2 \leq G$ for all t (same as convex, G-Lipschitz)
- Can show that there is a choice of η_t so that $f(\bar{x}_T) - \inf_{x \in \chi} f(x) \leq O\left(\frac{G\|x_0 - x_*\|_2}{\sqrt{t}}\right)$

Lemma: $g_y \stackrel{\text{def}}{=} b + \boldsymbol{A}\operatorname{sign}(\mathbf{A}^\top y - b) \in \partial[-f_{\min}]\,(y)$.

Dual Problem?

- $\|g_y\|_2$ could as large as $\sqrt{m}$
- Idea:
  - $\|g_y\|_\infty \leq \|b\|_\infty + \|\boldsymbol{A}\operatorname{sign}(\mathbf{A}^\top y - b)\|_\infty$
    $\leq \|b\|_\infty + \|\boldsymbol{A}\|_\infty$

($f_{\min}$ is convex and $\|b\|_\infty + \|\boldsymbol{A}\|_\infty$-Lipschitz in ℓ_∞)

Mirror Descent

Proximal Progress Lemma

For $w = \text{prox}_z^r(g)$ and all $u \in \chi$

$$g^\top(w-u) \le V_z^r(u) - V_z^r(w) - V_w^r(u)$$

Divergence

- $r: \chi \to \mathbb{R}$ convex and differentiable
- $V_x^r(y) \stackrel{\text{def}}{=} r(y) - [r(x) + \nabla r(x)^\top(y-x)]$

Proximal Step

- $\text{prox}_x^r(g) \in \text{argmin}_{y\in\chi} g^\top y + V_x^r(y)$

Strong Convexity

- Differentiable r is μ-strongly convex with respect to $\|\cdot\|$ if for all $x, y \in \chi$

$$r(y) \ge r(x) + \nabla r(x)^\top(y-x) + \frac{\mu}{2}\|y-x\|^2$$

Mirror Descent If r is 1-strongly convex with respect to $\|\cdot\|$ and $x_{t+1} = \text{prox}_{x_t}^r(g_t)$ for all $t \in [T]$. Then, $\forall u \in \chi$

$$\sum_{t\in[T]} g_t^\top(x_t - u) \le V_{x_1}^r(u) - V_{x_{T+1}}(u) + \frac{1}{2}\sum_{t\in[T]}\|g_t\|_*^2$$

Proof:

- $g_t^\top(x_{t+1} - u) \le V_{x_t}^r(u) - V_{x_t}^r(x_{t+1}) - V_{x_{t+1}}^r(u)$
- $g_t^\top(x_{t+1} - x_t) - V_{x_t}^r(x_{t+1})$

$$\le \|g_t\|_*\|x_{t+1} - x_t\| - \frac{1}{2}\|x_{t+1} - x_t\|^2$$

$$\le \frac{1}{2}\|g_t\|_*^2$$

- $g_t^\top(x_t - u) \le V_{x_t}^r(u) - V_{x_{t+1}}^r(u) + \frac{1}{2}\|g_t\|_*^2$

When r is differentiable equivalent to formal definition of $r(t\cdot x + (1-t)\cdot y) \le t\cdot r(x) + (1-t)\cdot r(y) - \frac{\mu t(1-t)}{2}\|x-y\|^2$.

Mirror Descent

Proximal Prgress Lemma

For $w = \text{prox}_z^r(g)$ and all $u \in \chi$

$$g^\top(w-u) \le V_z^r(u) - V_z^r(w) - V_w^r(u)$$

Divergence

- $r: \chi \to \mathbb{R}$ convex and differentiable
- $V_x^r(y) \stackrel{\text{def}}{=} r(y) - [r(x) + \nabla r(x)^\top(y-x)]$

Proximal Step

- $\text{prox}_x^r(g) \in \text{argmin}_{y\in\chi} g^\top y + V_x^r(y)$

Strong Convexity

- Differentiable r is μ-strongly convex with respect to $\|\cdot\|$ if for all $x, y \in \chi$

$$r(y) \ge r(x) + \nabla r(x)^\top(y-x) + \frac{\mu}{2}\|y-x\|^2$$

Mirror Descent If r is 1-strongly convex with respect to $\|\cdot\|$ and $x_{t+1} = \text{prox}_{x_t}^r(g_t)$ for all $t \in [T]$. Then, $\forall u \in \chi$

$$\sum_{t\in[T]} g_t^\top(x_t - u) \le V_{x_1}^r(u) - V_{x_{T+1}}(u) + \frac{1}{2}\sum_{t\in[T]}\|g_t\|_*^2 .$$

Corollary Let $f: \chi \to \mathbb{R}$ with $x_* \in \min_{x\in\chi} f(x)$

- $x_{t+1} = \text{prox}_{x_t}^r(-\eta^{-1} g_t)$ where $g_t \in \partial f(x_t)$
- $V_{x_1}^r(x_*) \le D$, $\|g_t\|_* \le G$, and $\eta = \sqrt{\frac{TG^2}{2D}}$

$\Rightarrow f(\bar{x}_T) - f(x_*) \le \sqrt{\frac{2G^2D}{T}}$ for $\bar{x}_T = \frac{1}{T}\sum_{t\in[T]} x_t$

Proof: $f(x_t) \ge f(x_*) + g_t^\top(x_t - x_*)$

- $\eta^{-1}\sum_{t\in[T]}[f(x_t) - f(x_*)] \le D + \frac{TG^2}{2\eta^2}$
- $\frac{1}{T}\sum_{t\in[T]} f(x_t) - f(x_*) \le 2\eta D$

A Dual ϵ^{-2} algorithm

Dual Problem

- Simplex: $\Delta^m \stackrel{\text{def}}{=} \{x \in \mathbb{R}^m_{\geq 0} \mid \|x\|_1 = 1\}$
- $\min_{y \in \Delta^m} -f_{\min}(y) = b^\top y + \|A^\top y - b\|_1$

Entropy regularizer

- $r_{\text{ent}}: \Delta^m \to \mathbb{R}$ defined as $r_{\text{ent}}(y) = \sum_i y_i \log y_i$
- r_{ent} is 1-strongly convex with respect to ℓ_1
- $\max_{y \in \Delta^m} r_{\text{ent}}(y) - \min_{y \in \Delta^m} r_{\text{ent}}(y) = O(\log n)$

Proximal Step

- $\text{prox}_y^r(g) = \frac{y_i \cdot \exp(-g_i)}{\sum_{i \in [m]} y_i \cdot \exp(-g_i)}$

Theorem: there is is an algorithm which computes an ϵ-approximate dual solution in time

$$O(\epsilon^{-2}(\|b\|_\infty + \|A\|_\infty)^2 \log n).$$

Run algorithm with $x_1 = \text{argmin}_{y \in \Delta^m} r_{\text{ent}}(y) = \frac{1}{m}\vec{1}$

Mirror Descent If r is 1-strongly convex with respect to $\|\cdot\|$ and $x_{t+1} = \text{prox}_{x_t}^r(g_t)$ for all $t \in [T]$. Then, $\forall u \in \chi$

$$\sum_{t \in [T]} g_t^\top (x_t - u) \leq V_{x_1}^r(u) - V_{x_{T+1}}(u) + \frac{1}{2}\sum_{t \in [T]} \|g_t\|_*^2 .$$

Corollary Let $f: \chi \to \mathbb{R}$ with $x_* \in \min_{x \in \chi} f(x)$

- $x_{t+1} = \text{prox}_{x_t}^r(-\eta^{-1} g_t)$ where $g_t \in \partial f(x_{t+1})$
- $V_{x_1}^r(x_*) \leq D$, $\|g_t\|_* \leq G$, and $\eta = \sqrt{\frac{TG^2}{2D}}$

$\Rightarrow f(\bar{x}_T) - f(x_*) \leq \sqrt{\frac{2G^2 D}{T}}$ for $\bar{x}_T = \frac{1}{T}\sum_{t \in [T]} x_t$

Lemma: $g_y \stackrel{\text{def}}{=} b + A\text{sign}(\mathbf{A}^\top y - b) \in \partial[-f_{\min}](y)$ and

$$\|g_y\|_\infty \leq \|b\|_\infty + \|A\|_\infty$$

Warmup Algorithms

- Box: $B_\infty^n \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid \|x\|_\infty \le 1\}$
- Simplex: $\Delta^m \stackrel{\text{def}}{=} \{x \in \mathbb{R}_{\ge 0}^m \mid \|x\|_1 = 1\}$
- $\min_{x \in B_\infty^n} \max_{y \in \Delta^m} f(x,y) \stackrel{\text{def}}{=} y^\top \boldsymbol{A} x + c^\top x - b^\top y$

Primal Problem

- $\min_{x \in B_\infty^n} f_{\max}(x)$
- $f_{\max}(x) = \max_{y \in \Delta^m} f(x,y)$
- $f_{\max}(x) = c^\top x + \max_{i \in [m]} [\boldsymbol{A}x - b]_i$

ℓ_∞-Gradient Descent

- $\sim\epsilon^{-2}$ iteration method

Dual Problem

- $\max_{y \in \Delta^m} f_{\min}(y)$
- $f_{\min}(y) = \min_{x \in B_\infty^n} f(x,y)$
- $f_{\min}(y) = -b^\top y - \|\boldsymbol{A}^\top y - b\|_1$

Mirror Descent in ℓ_1

- $\sim\epsilon^{-2}$ iteration method

How to improve?

Primal dual method!

Primal Dual

- Box: $B_\infty^n \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid \|x\|_\infty \leq 1\}$
- Simplex: $\Delta^m \stackrel{\text{def}}{=} \{y \in \mathbb{R}_{\geq 0}^m \mid \|y\|_1 = 1$
- $\min_{x \in B_\infty^n} \max_{y \in \Delta^m} f(x,y) \stackrel{\text{def}}{=} y^\top \boldsymbol{A} x + c^\top x - b^\top y$
- $\text{gap}(x,y) = f_{\max}(x) - f_{\min}(y)$

Notation

- $z \in B_\infty^n \times \Delta^m$ with $z = (z^x, z^y)$
- $z^x \in B_\infty^n$ and $z^y \in \Delta^m$

Approach

- $g(z) \stackrel{\text{def}}{=} (\nabla_x f(z), -\nabla_y f(z))$
- $\nabla_x f(z) = \boldsymbol{A}^\top z^y + c$
- $\nabla_y f(z) = \boldsymbol{A} z^x - b$

Bound

- Suppose $\frac{1}{T}\sum_{t \in [T]} g(z_t)^\top (z_t - u) \leq \epsilon$
- For $\bar{z} = \frac{1}{T}\sum_{t \in [T]} z_t$ have $\text{gap}(\bar{z}^x, \bar{z}^y) \leq \epsilon$

Mirror Descent?

- Idea: apply to $g(z_t)$

Problem #1

- Want ϵ^{-1} rate
- Idea: mirror prox! smoothness!

Problem #2

- Strongly convex r on $B_\infty^n \times \Delta^m$ with $\sup_z r(z) - \inf_z r(z)$ bounded?
- **Thm**: any r that is 1-strongly convex with respect to $\|\cdot\|_\infty$ has $\sup_z r(z) - \inf_z r(z) = \Omega(n)$
- Ideas:
  - r can have interaction between B_∞^n and Δ^n
  - Closer analysis of relation between r and g

Thank you

Questions?

Aaron Sidford

Contact Info:

- email: *sidford@stanford.edu*
- website: *www.aaronsidford.com*