Optimization and Randomization Approaches to Games and Teams

Ankur A. Kulkarni

Systems and Control Engineering
Indian Institute of Technology Bombay

kulkarni.ankur@iitb.ac.in

Spans joint work with Todd Coleman (UCSD), Bharat Prabhakar and Sharu Theresa Jose

Overview

	Problems	
	Deterministic problems from **game theory**	Stochastic problems from **team theory**
Applications	Electricity markets, transportation networks etc	Decentralized control theory, information theory
Nature	Deterministic	Stochastic
Interest	Algorithms	Bounds, fundamental limits
Approach	Randomization	Optimization, de-randomization

- Interplay between convex analytic geometry and probability

Fast approximations to high-dimensional AVIs and LCPs

- Vast amounts of data is being created and stored
- **High dimensional problems** $\Longrightarrow$ exact algorithms expensive
- **Speed v/s accuracy tradeoff**
- Need for quick approximations, even if at the expense of accuracy

Standard approach

- Solve the given problem, but to lesser accuracy
- Use calculus or relaxations to construct an approximate problem

Contribution

- A new method for obtaining fast approximation

Problem class: LCPs and AVIs

Linear Complementarity Problem: $\mathrm{LCP}(M, q)$

Given $M \in \mathbb{R}^{n \times n}, q \in \mathbb{R}^n$, find x such that

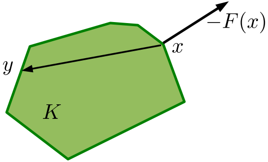
$$0 \leq x \perp Mx + q \geq 0.$$

Affine Variational Inequality: $\mathrm{AVI}(K, M, q)$

Given a **polyhedron** $K \subseteq \mathbb{R}^n$ and an affine mapping $F(x) \equiv q + Mx$, $\mathrm{AVI}(K, M, q)$, is to find a vector $x \in K$ such that

$$(y - x)^\top (Mx + q) \geq 0, \quad \forall y \in K.$$

- $\mathrm{LCP}(M, q) \equiv \mathrm{AVI}(\mathbb{R}^n_+, M, q)$
- Known to be **NP-hard**
- Includes QPs, equilibria of games etc



Problem Definition (compact K)

Approximate solution $\mathrm{AVI}(K, M, q)$

x is an $\hat{\epsilon}$-approximate solution, $\hat{\epsilon} = \hat{\epsilon}(\textit{problem constants}) > 0$, if

$$(y - x)^T (q + Mx) \geq -\hat{\epsilon}, \ \forall y \in K$$

Desiderata: We seek an **approximation algorithm** that satisfies the following requirements

- The algorithm must involve solving a lower (preferably $o(n)$) dimensional problem
- The lower dimensional problem should be **also be an AVI** (or some canonical problem)
- The **guarantees** for the algorithm **need not be deterministic**

Approach overview

- Given $\mathrm{AVI}(K, M, q)$
- **Randomly "project" (suitably)** it to create a random lower dimensional $\mathrm{AVI}(\widetilde{K}, \widetilde{M}, \widetilde{q})$
- Solve the lower dimensional problem $\longrightarrow \widetilde{x}$
- Obtain a random approximation to the solution of the given AVI

Main claim

- The **(deterministic) preimage** (under the random projection) of $\widetilde{x}$ approximately solves the given AVI.

Main tool

- Johnson-Lindenstrauss lemma

The Johnson-Lindenstrauss Lemma

Lemma (Preservation of Multiple Norms)

Let $R \in \mathbb{R}^{n \times k}$ be a uniformly random orthonormal matrix, and let $f(u) = \sqrt{\frac{n}{k}} R^\top u$ for $u \in \mathbb{R}^n$. Then for any $0 < \epsilon < 1$ and $u_1, \cdots, u_m$,

$$P(|\|u_i\|^2 - \|f(u_i)\|^2| \leq \epsilon \|u_i\|^2 \quad \forall i = 1, \cdots, m) \geq 1 - 2me^{-(\epsilon^2/2 - \epsilon^3/3)k/2}$$

The Johnson-Lindenstrauss Lemma

Lemma (Preservation of Multiple Norms)

Let $R \in \mathbb{R}^{n\times k}$ *be a uniformly random orthonormal matrix, and let* $f(u) = \sqrt{\frac{n}{k}} R^\top u$ *for* $u \in \mathbb{R}^n$*. Then for any* $0 < \epsilon < 1$ *and* $u_1, \cdots, u_m$*,*

$$P(|\|u_i\|^2 - \|f(u_i)\|^2| \le \epsilon \|u_i\|^2 \quad \forall i = 1, \cdots, m) \ge 1 - 2me^{-(\epsilon^2/2 - \epsilon^3/3)k/2}$$

Lemma (Preservation of Multiple Inner Products)

Let $u_i, v_i, i = 1, \cdots, m$ *be a set of points in* $\mathbb{R}^n$*. For* $0 < \epsilon < 1$*,*

$$P\left(|u_i^\top v_i - f(u_i)^\top f(v_i)| \le \epsilon \|u_i\| \|v_i\| \ \forall i = 1, \cdots, m\right) \ge 1 - 4me^{-(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})k/2}$$

The Johnson-Lindenstrauss Lemma

Lemma (Preservation of Multiple Norms)

Let $R \in \mathbb{R}^{n\times k}$ be a uniformly random orthonormal matrix, and let $f(u) = \sqrt{\frac{n}{k}} R^\top u$ for $u \in \mathbb{R}^n$. Then for any $0 < \epsilon < 1$ and $u_1, \cdots, u_m$,

$$P(|\|u_i\|^2 - \|f(u_i)\|^2| \le \epsilon \|u_i\|^2 \quad \forall i = 1, \cdots, m) \ge 1 - 2me^{-(\epsilon^2/2 - \epsilon^3/3)k/2}$$

Lemma (Preservation of Multiple Inner Products)

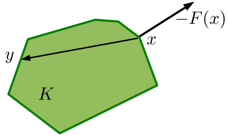
Let $u_i, v_i, i = 1, \cdots, m$ be a set of points in $\mathbb{R}^n$. For $0 < \epsilon < 1$,

$$P\left(|u_i^\top v_i - f(u_i)^\top f(v_i)| \le \epsilon \|u_i\| \|v_i\| \ \forall i = 1, \cdots, m\right) \ge 1 - 4me^{-(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})k/2}$$

- $1 - 4me^{-(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})k/2} \ge 1 - \delta \iff k \approx \frac{1}{\epsilon^2} \ln m/\delta$.
- Only **finitely many** inner products allowed
- The dimension n appears nowhere; instead m appears
- ϵ can be chosen independently of n, m
- Notice the scaling/normalization

Key idea

- Solving an AVI amounts to finding an x s.t. $Mx + q$ makes certain angles/inner products with $y - x$ as y ranges over K



-

$$P(|\underbrace{u_i^\top v_i}_{=\mathbb{E}[f(u_i)^\top f(v_i)]} - f(u_i)^\top f(v_i)| \leq \epsilon \|u_i\| \|v_i\| \;\; \forall i = 1, \cdots, m) \geq 1 - 4me^{-(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})k/2}$$

- **General idea:** Craft a lower dimensional problem such that:

  $u_i^\top v_i \approx$ optimality for higher dimensional problem

  $f(u_i)^\top f(v_i) \approx$ optimality for lower dimensional problem

- **But there is a catch!** – does not lead to an AVI as a lower-dimensional problem
- Need a **correction, adds an error**; need careful crafting to minimize error

Challenges

Convex analysis

- Simultaneously minimize the lower dimension and error
- AVIs involve **infinitely many** inner products; need to **reduce to finitely many**
- Make sure they are **few enough** – exploit the convex geometry

Solution recovery

Challenges

Convex analysis

- Simultaneously minimize the lower dimension and error
- AVIs involve **infinitely many** inner products; need to **reduce to finitely many**
- Make sure they are **few enough** – exploit the convex geometry

Solution recovery

- How to **back project** from the lower dimension?
- **Back projection** error? How to drive this to zero?

Challenges

Convex analysis

- Simultaneously minimize the lower dimension and error
- AVIs involve **infinitely many** inner products; need to **reduce to finitely many**
- Make sure they are **few enough** – exploit the convex geometry

Solution recovery

- How to **back project** from the lower dimension?
- **Back projection** error? How to drive this to zero?
- Will there be **tradeoffs** or can we achieve everything?

Challenges

Convex analysis

- Simultaneously minimize the lower dimension and error
- AVIs involve **infinitely many** inner products; need to **reduce to finitely many**
- Make sure they are **few enough** – exploit the convex geometry

Solution recovery

- How to **back project** from the lower dimension?
- **Back projection** error? How to drive this to zero?
- Will there be **tradeoffs** or can we achieve everything?
- **What we would like:** for fixed problem constants, we can compute tolerance. For fixed tolerance, lower dimension must be $o(n)$. As $n \to \infty$ error is within tolerance. In short no tradeoffs.

Algorithm (Compact K) I

1. Fix the error parameter $\epsilon \in (0,1)$ and the success confidence parameter $\delta \in [0,1)$. Pick the lower dimension value k such that $k \geq \frac{2\ln(\frac{4\eta}{\delta})}{(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})}$, where $\eta = |ext(K)|$.
2. Construct $n \times k$ random orthonormal matrix R
3. Construct lower dimensional $\mathrm{AVI}(\widetilde{K}, \widetilde{M}, \widetilde{q})$. Define

$$\widetilde{q} = \sqrt{\frac{n}{k}} R^\top q \quad \widetilde{M} = cR^\top MR \quad \widetilde{K} = \left\{ \widetilde{x} \in \mathbb{R}^k | \widetilde{x} = \sqrt{\frac{n}{k}} R^\top x, Ax \leq b \right\}$$

4. Solve $\mathrm{AVI}(\widetilde{K}, \widetilde{M}, \widetilde{q}) \longrightarrow \widetilde{x}$.
5. Solve the L_1 minimization problem

$$\min_{x \in \mathbb{R}^n} \quad \|x\|_1 \qquad \longrightarrow \qquad x^*$$
$$\text{s.t.} \quad \sqrt{\frac{n}{k}} R^\top x = \widetilde{x}$$

6. Project x^* on K,

$$x^\# = \underset{x \in K}{\arg\min} \quad \|x^* - x\|^2$$

7. $x^\#$ is the final output vector.

Analysis

$\widetilde{x} \longleftarrow \mathrm{AVI}(\widetilde{K}, \widetilde{M}, \widetilde{q})$
Let x_o be a deterministic pre-image such that $\sqrt{\frac{n}{k}} R^\top x_o = \widetilde{x}$.

$$(\widetilde{y} - \widetilde{x})^\top (\widetilde{q} + \widetilde{M}\widetilde{x}) \geq 0, \quad \forall \widetilde{y} \in \widetilde{K} \quad (\textit{infinitely many})$$

$$\implies \sqrt{\frac{n}{k}} (R^\top (y - x_o))^\top \sqrt{\frac{n}{k}} R^\top (q + M' x_o) \geq 0, \quad \forall y \in K \quad (M' = cMRR^\top)$$

$$\implies \sqrt{\frac{n}{k}} (R^\top (x_e - x_o))^\top \sqrt{\frac{n}{k}} R^\top (q + M' x_o) \geq 0, \quad \forall x_e \in ext(K) \quad (\textit{finite!})$$

$M' = cMRR^\top$. Preserving $\eta = |ext(K)|$ inner products, **JL lemma** $\implies$ w.p $\geq (1 - 4\eta e^{-(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})k/2})$, we have

$$(x_e - x_o)^\top (q + Mx_o) \geq \frac{n}{k} (x_e - x_o)^\top RR^\top (M - M') x_o + (-\epsilon \|x_e - x_o\| \|q + Mx_o\|) \quad \forall x_e \in ext(K).$$

Error Bounds

Since the error depends on $(M - M')$ it is tempting to take $M' = MRR^\top$ (i.e., $c = 1$).

some lines of bounding arguments later

$$\implies (y - x_o)^\top (q + Mx_o) \geq -\frac{\left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}\right)n}{\ln\left(\frac{4\eta}{\delta}\right)} \|M\| \|x_o\| - \epsilon \cdot D\|q\| - \epsilon \cdot D\|M\| \|x_o\|, \ \forall y \in K$$

Remember:

- There are several random x's that satisfy $\sqrt{\frac{n}{k}} R^\top x = \widetilde{x}$.
- The above result pertains to only (and all) $x = x_o$'s that are deterministic.

Finding x_o

- The algorithm gives us $\widetilde{x}$ (solution of lower dim problem)
- Need to backproject to find a **deterministic** $x_o \in \mathbb{R}^n$ from $\widetilde{x} \in \mathbb{R}^k$
- $\implies$ Problem from *compressed sensing*: recovering a "signal" from its "partial observations"
- Many results which provide "high probability" guarantees on **exact** recovery of the sparsest x_o. The result below is requires no sparsity.

Theorem ([CT06])

Suppose that $f \in \mathbb{R}^n$ obeys $\|f\|_1 \leq C_1$, and we are given k uniformly random orthonormal measurements $y = \sqrt{\frac{n}{k}} R^\top f$. Then with probability 1, we have a unique minimizer $f^\#$ to the following problem:

$$\min_{x \in \mathbb{R}^n} \quad \|x\|_1 \qquad \text{such that } y = \sqrt{\frac{n}{k}} R^\top x \quad .$$

Furthermore, with probability at least $1 - O(n^{-1/\alpha})$, we have the approximation

$$\|f - f^\#\| \leq C \cdot C_1 \cdot (\ln(n)/k)^{1/2}.$$

Here, C is a fixed constant depending on α but not on anything else. The implicit constant in $O(n^{-1/\alpha})$ is allowed to depend on α.

First result ($c = 1$) [PK14a] [PK14b]

Theorem

1. $\widetilde{x}$ *exists w.p. 1*
2. *For every* $\epsilon \in (0,1)$, *if the lower dimension* k *satisfies* $k \geq \frac{2\ln(\frac{4\eta}{\delta})}{(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})}$, $0 \leq \delta < 1$, *then with a probability of at least* $(1-\delta)$, *the deterministic pre-image* x_o *satisfies,*

$$(y - x_o)^\top (q + Mx_o) \geq \hat{\epsilon}, \ \forall y \in K$$

where $\hat{\epsilon} = -\frac{(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})n}{2ln(\frac{4\eta}{\delta})} \|M\| \|x_o\| - \epsilon \cdot D\|q\|_2 - \epsilon \cdot D\|M\| \|x_o\|.$

3. *With a probability at least* $1 - O(n^{-1/\alpha})$:

$$\|x_o - x^{\#}\| \leq C \cdot \|x_o\|_1 \cdot (ln(n)/k)^{1/2}$$

where $\alpha > 0$ *is a sufficiently small number and* C *is a constant depending only on* α.

Revisiting the bound

- Lower dimension = $O(\ln \eta)$ where $\eta = |ext(K)|$
- **Good:** reveals something combinatorial about the problem
- **Bad:** requires K to be a polytope, does not cover LCPs
- **Bad:** worst case $\ln \eta \approx n$.
- **Unclear:** is this the true complexity?
- **Error:** Can the error be made smaller?

Revisiting the bound

- Taking $c = 1 \iff \widetilde{M} = R^\top MR \iff$ pseudoinverse – **Too conservative.**
- Turns out: $c = \frac{n}{k}$ can give **lower error**, but with needs preservation of **more inner products**
- But not too many more!
- **Key idea:** Matrix-vector inner product preservation

$$P\left(\| \frac{n}{k} d^\top RR^\top A - d^\top A \| \leq \epsilon \|d\| \|A\|_F \right) \geq 1 - 4ne^{-(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})k/2},$$

Improved result

Theorem

Let $c = \frac{n}{k}$. For every $\epsilon \in (0,1)$ and $0 < \delta \leq 1$, if the lower dimension k satisfies $k \geq 2\ln(\frac{4(n+n\eta+\eta)}{\delta})/(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})$, then with a probability of at least $(1-\delta)$, the deterministic vector x_o solves $\mathrm{AVI}(q,\ M,\ K)$ approximately, i.e.,

$$(y - x_o)^\top (q + Mx_o) \geq -\hat{\epsilon},\ \ \forall y \in K$$

where $\hat{\epsilon} = \epsilon\sqrt{1 + 2\epsilon + \epsilon^2 n}\|M\|_F\|x_o\| + \epsilon \cdot D\|q\|_2 + \epsilon \cdot D\|M\|\|x_o\|$.

Improved result

Theorem

Let $c = \frac{n}{k}$. For every $\epsilon \in (0,1)$ and $0 < \delta \leq 1$, if the lower dimension k satisfies $k \geq 2\ln(\frac{4(n+n\eta+\eta)}{\delta})/(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})$, then with a probability of at least $(1-\delta)$, the deterministic vector x_o solves $\mathrm{AVI}(q, M, K)$ approximately, i.e.,

$$(y - x_o)^\top (q + Mx_o) \geq -\hat{\epsilon}, \;\; \forall y \in K$$

where $\hat{\epsilon} = \epsilon\sqrt{1 + 2\epsilon + \epsilon^2 n}\|M\|_F\|x_o\| + \epsilon \cdot D\|q\|_2 + \epsilon \cdot D\|M\|\|x_o\|$.

Corollary

With $\epsilon = \frac{\epsilon'}{n^{\frac{1}{4}}}$, we get $\hat{\epsilon} = -\epsilon' \cdot const(M, q, K)$ and $k \sim \frac{1}{(\epsilon')^2}\sqrt{n}\ln(\frac{4(n+n\eta+\eta)}{\delta})$.

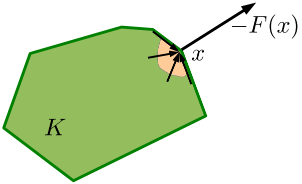
- ϵ' can be selected based on only the constants in the problem
- $k = o(n)$ and $\frac{\ln n}{k} \to 0$. Thus back projection error also $\to 0$

Further improvement [PK15]

- How large is $\eta = |ext(K)|$? If $K = \{x | Ax = b, x \geq 0\}$ and $A \in \mathbb{R}^{m \times n}$ is full row rank, $\eta \leq \binom{n}{m}$ In other representations, there is no easy bound.
- Analysis required compactness of K. What if K is unbounded?

Tangent cone

$$\mathcal{T}(x;K) = \left\{ d \mid \exists \{x_k\} \subseteq K, x_k \to x, \tau_k \downarrow 0, \text{ such that } d = \lim_{k\to\infty} \frac{x_k - x}{\tau_k} \right\}.$$



$$x \text{ solves AVI}(K, M, q) \iff (Mx+q)^\top d \geq 0 \qquad \forall d \in \underbrace{\mathcal{T}(x;K)}_{\textit{again infinite!}}.$$

Further improvement

Suffices to preserve inner products with only the generators of the tangent cone

Lemma

If K is polyhedral, $\mathcal{T}(x;K)$ is polyhedral. A point $x \in K$ solves $\mathrm{AVI}(K, M, q)$ if and only if

$$d^\top F(x) \geq 0, \qquad \forall d \in conv\left(ext(\mathcal{T}(x;K))\right).$$

Further improvement

Suffices to preserve inner products with only the generators of the tangent cone

Lemma

If K is polyhedral, $\mathcal{T}(x;K)$ is polyhedral. A point $x \in K$ solves $\mathrm{AVI}(K,M,q)$ if and only if

$$d^\top F(x) \geq 0, \qquad \forall d \in conv\left(ext(\mathcal{T}(x;K))\right).$$

Lemma

Let K be a polyhedron, $x \in K, \widetilde{x} = \sqrt{\frac{n}{k}} R^\top x$ and let $\widetilde{K} = \sqrt{\frac{n}{k}} R^\top K$. Then,

$$\sqrt{\frac{n}{k}} R^\top \mathcal{T}(x;K) = \mathcal{T}(\widetilde{x};\widetilde{K}).$$

- Proceed as before ...

$$(\tilde{M}\tilde{x} + \tilde{q})^\top \widetilde{d} \geq 0 \qquad \forall \widetilde{d} \in \underbrace{\mathcal{T}(\widetilde{x};\widetilde{K})}_{\sqrt{\frac{n}{k}} R^\top \mathcal{T}(x_o;K) \implies \sqrt{\frac{n}{k}} R^\top ext(\mathcal{T}(x_o;K))}$$

Applying to LCPs

High dim LCPs → low dimensional AVI

Theorem

Let $\epsilon \in (0,1), \delta \in (0,1]$ *and* $\mathrm{LCP}(M,q)$ *be provided as input to Algorithm. If the lower dimension* k *satisfies* $k \geq 2\ln\left(\frac{4}{\delta}(n^2+2n)\right)/\left(\frac{\epsilon^2}{2}-\frac{\epsilon^3}{3}\right)$, *then* **conditioned on the event that** $\widetilde{x}$ **exists** *with probability at least* $(1-\delta)$, x_o *satisfies*

$$|(q+Mx_o)_i| \leq \epsilon\|q+Mx_o\| + \epsilon\sqrt{1+\epsilon^2 n+2\epsilon}\|M\|_F\|x_o\| \quad \forall i \notin \mathcal{A}(x_o)$$

$$(q+Mx_o)_i \geq -\epsilon\|q+Mx_o\| - \epsilon\sqrt{1+\epsilon^2 n+2\epsilon}\|M\|_F\|x_o\| \quad \forall i \in \mathcal{A}(x_o)$$

$$|x_o^\top(q+Mx_o)| \leq \epsilon\|x_o\|_1\|q+Mx_o\| + \epsilon\sqrt{1+\epsilon^2 n+2\epsilon}\|x_o\|_1\|M\|_F\|x_o\|$$

where $\mathcal{A}(x) = \{i | x_i = 0\}$

Applying to LCPs

High dim LCPs → low dimensional AVI

Theorem

Let $\epsilon \in (0,1), \delta \in (0,1]$ *and* $\mathrm{LCP}(M,q)$ *be provided as input to Algorithm. If the lower dimension* k *satisfies* $k \geq 2\ln\left(\frac{4}{\delta}(n^2+2n)\right)/\left(\frac{\epsilon^2}{2}-\frac{\epsilon^3}{3}\right)$, *then* **conditioned on the event that** $\widetilde{x}$ **exists** *with probability at least* $(1-\delta)$, x_o *satisfies*

$$|(q+Mx_o)_i| \leq \epsilon\|q+Mx_o\| + \epsilon\sqrt{1+\epsilon^2 n+2\epsilon}\|M\|_F\|x_o\| \quad \forall i \notin \mathcal{A}(x_o)$$

$$(q+Mx_o)_i \geq -\epsilon\|q+Mx_o\| - \epsilon\sqrt{1+\epsilon^2 n+2\epsilon}\|M\|_F\|x_o\| \quad \forall i \in \mathcal{A}(x_o)$$

$$|x_o^\top(q+Mx_o)| \leq \epsilon\|x_o\|_1\|q+Mx_o\| + \epsilon\sqrt{1+\epsilon^2 n+2\epsilon}\|x_o\|_1\|M\|_F\|x_o\|$$

where $\mathcal{A}(x) = \{i | x_i = 0\}$

- **Conditional probability:** Lower dim AVI obtained from LCPs need not be solvable

Corollary

Can take $k \sim \frac{1}{(\epsilon')^2}\sqrt{n}\ln(n^2+2n) = o(n)$. *Back projection error* $\to 0$

AVIs on noncompact polyhedra

Theorem

Let $\epsilon, \delta \in (0,1)$ *and suppose* $\mathrm{AVI}(K, M, q)$ *is provided as input to the algorithm and let* $k \geq 2\ln\left(\frac{4}{\delta}(n + n\kappa + \kappa)\right)/(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})$, *where*

$$\kappa = \max_{x \in K} |ext(\mathcal{T}(x; K))| = O(n)^*.$$

1. *If* $\mathrm{AVI}(K, M, q)$ *has a nondegenerate solution,* $\widetilde{x}$ *exists with high probability*
2. *Conditioned on the event that* $\widetilde{x}$ *exists, with probability greater than* $1 - \delta$, x_o *satisfies,*

$$d_e^\top (q + Mx_o) \geq -\epsilon \cdot \sqrt{1 + 2\epsilon + \epsilon^2 n} \|M\|_F \|x_o\| - \epsilon \cdot \|q + Mx_o\| \qquad \forall\ d_e \in ext(\mathcal{T}(x_o; K)).$$

- $^*ext(\bullet)$ = set of generators of $\bullet$. For a polytope K, $|ext(\mathcal{T}(x; K))| \geq |ext(K)|$. Under a suitable representation for K, $\kappa = O(n)$

Corollary

Once again, we may take $k \sim \frac{1}{(\epsilon')^2}\sqrt{n}\ln(n + n\kappa + \kappa)$.

Further remarks

Solving the lower dimensional problem

- Lower dim problem has implicit constraints
- Make implicit variables explicit – sparse problem higher dimensional problem
- Need a solver that can exploit constraints like these

Numerical performance

- Much faster; accuracy improves as k increases
- Useful for quickly deriving an initial point
- Substantial savings above random starting points

Approximate solutions for LCPs

n	k	Major (low)	Minor (low)	Major (direct)	Minor (direct)	Natural Map Residual	Inner Prod	Max Inner Prod	Min Inner Prod	ϵ'
400	50	4	36	1	342	0.95	0.57	1.05	0.00	0.005
400	60	7	233	1	342	0.95	10.04	8.91	0.00	0.005
400	70	4	19	1	342	0.96	1.22	1.44	0.00	0.006
500	70	4	25	1	196	0.93	0.98	1.45	0.00	0.004
500	80	4	4	1	196	0.98	1.21	1.25	0.00	0.005
500	90	5	13	1	196	0.96	1.30	1.31	0.00	0.005
600	70	2	3	1	111	0.98	0.39	0.51	0.00	0.004
600	80	3	6	1	111	0.94	0.79	1.07	0.00	0.004
600	90	7	11	1	111	0.97	1.51	1.93	0.00	0.004
700	75	4	22	1	655	0.94	0.33	0.45	0.00	0.003
700	100	5	28	1	655	0.96	0.70	0.63	0.00	0.004
700	125	14	86	1	655	0.96	2.14	1.35	0.00	0.004
700	150	14	348	1	655	0.98	18.58	10.49	0.00	0.004
850	100	4	419	1	166	0.95	5.41	6.01	0.00	0.003
850	125	9	1010	1	166	0.97	1.86	1.40	0.00	0.003
850	150	11	121	1	166	0.98	2.71	1.82	0.00	0.004
1000	100	6	245	1	1269	0.97	2.79	3.59	0.00	0.002
1000	125	6	141	1	1269	0.96	3.17	2.32	0.00	0.003
1000	150	13	1109	1	1269	0.97	0.90	0.72	0.00	0.003

Exact solutions of AVIs

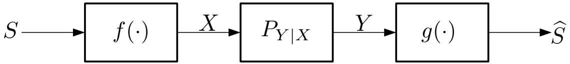
n	k	m	Major (New)	Major (Old)	Major Total	Minor (New)	Minor (Old)	Minor Total	Time (new)	Time (direct)
300	10	10	0	3	3	0	3156	13	0.07	0.33
300	50	10	1	3	5	312	3156	328	0.11	0.33
300	90	10	1	3	6	622	3156	639	0.12	0.33
300	130	10	1	3	7	312	3156	329	0.11	0.33
400	25	10	1	2	5	4	4126	18	0.19	0.55
400	50	10	1	2	5	418	4126	432	0.21	0.55
400	100	10	1	2	7	834	4126	851	0.25	0.55
400	150	10	1	2	8	834	4126	852	0.26	0.55
400	200	10	1	2	8	1246	4126	1265	0.28	0.55
600	25	10	5	46	9	8225	58025	8239	2.44	19.27
600	50	10	7	46	11	10178	58025	10193	3.03	19.27
600	100	10	6	46	12	8888	58025	8904	2.68	19.27
600	200	10	6	46	13	10155	58025	10172	2.98	19.27
800	25	10	0	2	4	1	8137	15	0.65	2.52
800	50	10	0	2	5	820	8137	835	0.81	2.52
800	100	10	1	2	6	2	8137	18	0.67	2.52
800	300	10	1	2	8	1634	8137	1652	0.99	2.52
1000	50	10	0	3	5	2	10130	17	1.14	5.86
1000	100	10	1	3	6	1011	10130	1027	1.51	5.86
1000	300	10	1	3	9	1006	10130	1039	1.44	5.86

Team decision problems

Team decision problems

- Random environment (prerealized but unknown) $\xi \in \mathbb{R}^n$
- N players
- Players get information, denoted $z_1, \cdots, z_N$; $z_i \in \mathbb{R}^{m_i}$
- They have to choose "strategies", i.e., measurable functions $\gamma_i : \mathbb{R}^{m_i} \to \mathbb{R}$ to jointly minimize
$$J(\gamma_1, \cdots, \gamma_N) = \mathbb{E}[L(u_1, \cdots, u_N; \xi)],$$
where $u_i = \gamma_i(z_i)$ and $L : \mathbb{R}^{N+n} \to \mathbb{R}$
- Complications arise due to z_i
- **Simplest case** is when $z_i = \eta_i(\xi)$. Studied first by [Rad62].
- **Hard case** is when $z_i = \eta_i(u_{S_i}, \xi)$; even simple instances are unsolved [Wit68].
Solvable subcase "classical" stochastic control. There is a time axis and
$$S_i = u_1, \cdots, u_{i-1} \quad \text{and } z_i \supseteq z_{i-1}$$
- Without this, extremely hard. Throw up fundamental issues such causality, "who knows what" etc

Introduction

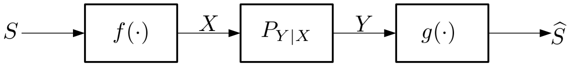


Stochastic optimal control problem with *non-classical information pattern*:

$$\text{SC} \quad \underset{f,g}{\text{minimize}} \quad \mathbb{E}[\kappa(S,X,Y,\widehat{S})]$$
$$X = f(S)$$
$$\widehat{S} = g(Y).$$

- $S, X, Y, \widehat{S}$ take values in $\mathcal{S}, \mathcal{X}, \mathcal{Y}, \widehat{\mathcal{S}}$.
- **Given:** $S \sim P_S,\ Y|X \sim P_{Y|X}$.
- **Information structure:** $S \to X \to Y \to \widehat{S}$
- **Cost:** $\kappa : \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \widehat{\mathcal{S}} \to \mathbb{R}$
- **Decision variables:** $f : \mathcal{S} \to \mathcal{X},\ g : \mathcal{Y} \to \widehat{\mathcal{S}}$

Examples and challenges



- **Witsenhausen:**

$$Y = X + W, \quad S, W \sim \text{Gaussian}, \quad \kappa(S, X, Y, \widehat{S}) \equiv (S - X)^2 + k(\widehat{S} - X)^2$$

- **Finite block-length channel coding:**

$$S \sim \text{Unif}(\mathcal{S}), \quad P_{Y|X} = \prod_{i=1}^{n} P_{Y_i|X_i}, \quad \kappa(S, X, Y, \widehat{S}) \equiv \mathbb{I}_{\{S \neq \widehat{S}\}}$$

- **Finite block-length lossy source-channel coding:**

$$\mathcal{S} = \prod_{i=1}^{n} \mathcal{S}_i \quad P_{Y|X} = \prod_{i=1}^{n} P_{Y_i|X_i}, \quad \kappa(S, X, Y, \widehat{S}) \equiv \frac{1}{n} \sum_{i=1}^{n} d(S_i, \widehat{S}_i)$$

- **Common fundamental challenge: Obtaining lower bounds**

Standard approach

- A block of n symbols is encoded as $(X_1, \cdots, X_n) = f(S_1, \cdots, S_n)$; this entails a natural delay
- Embed the given finite blocklength problem in an "infinite" blocklength version (equivalent to allowing arbitrary delay)
- Use results from information theory to solve the infinite blocklength problem, which bounds the given problem
- This leads to bounds that **necessarily involve** mutual information.

Lower bounds via convex relaxation

[KC12, KC15]

Original problem

$$\underset{f,g}{\text{minimize}} \quad \mathbb{E}[\kappa(S,X,Y,\widehat{S})]$$
$$X = f(S)$$
$$\widehat{S} = g(Y).$$

$\Longleftrightarrow$

Optimization over distributions

(Nonconvex!)

$$\underset{Q}{\text{minimize}} \quad \mathbb{E}[\kappa(S,X,Y,\widehat{S})]$$
$$Q_S = P_S, Q_{Y|X} = P_{Y|X}$$
$$S \to X \to Y \to \widehat{S}.$$

$\Downarrow$

A particular relaxation

$$\underset{Q}{\text{minimize}} \quad \mathbb{E}[\kappa(S,X,Y,\widehat{S})]$$
$$Q_S = P_S, \; Q_{Y|X} = P_{Y|X}$$
$$(DPI) \; I(S;\widehat{S}) \le I(X;Y)$$

$\Longleftarrow$

Convex relaxation

$$\underset{Q}{\text{minimize}} \quad \mathbb{E}[\kappa(S,X,Y,\widehat{S})]$$
$$Q \in \text{larger convex set}$$

Consequences:

- Infinite blocklength bounds are the **same** as those from convex relaxation by mutual information. This relaxation is **tight** for the variant of Witsenhausen in [BB87]
- New proofs, generalizations etc.

What is a good relaxation?

[JK15, JK16]

- **Ideal relaxation** – convex hull!
- Solutions of the original problem and of the relaxation lie on extreme points of respective feasible regions
- A good relaxation must *include all extreme points* of the original problem *as extreme points* of its feasible region
- Simplicity – solving/bounding the relaxed problem should not itself be hard

Shortcomings of the DPI relaxation

- DPI relaxation **does not have this property**
- $\implies$ There are costs for which the DPI relaxation is not tight

This talk

- A linear programming relaxation with the required property
- Show that it is tight in certain cases, derive new bounds

Linear programming relaxation

- $\mathcal{S}, \mathcal{X}, \mathcal{Y}, \widehat{\mathcal{S}}$ are finite
- Non-convexity is due to bilinear terms "$Q_{X|S}(x|s)Q_{\widehat{S}|Y}(\widehat{s}|y)$"

Relaxation:

1. Replace bilinear terms with new variables W:

$$W(s,x,y,\widehat{s}) = Q_{X|S}(x|s)Q_{\widehat{S}|Y}(\widehat{s}|y)$$

2. Add linear constraints that $W(s,x,y,\widehat{s})$ must satisfy, e.g.,

$$\sum_x W(s,x,y,\widehat{s}) = Q_{\widehat{S}|Y}(\widehat{s}|y), \qquad \forall \widehat{s}, y$$

$$\text{LP} \min_{Q_{X|S}, Q_{\widehat{S}|Y}, W} \quad \sum_z \kappa(z) P_S(s) P_{Y|X}(y|x) W(s,x,y,\widehat{s})$$

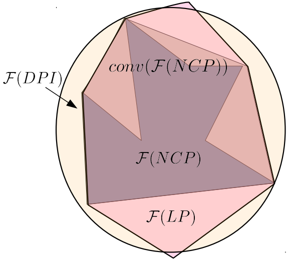
$$\text{subject to} \quad \begin{aligned}
\textstyle\sum_x Q_{X|S}(x|s) &= 1 && \forall s \\
\textstyle\sum_{\widehat{s}} Q_{\widehat{S}|Y}(\widehat{s}|y) &= 1 && \forall y \\
\textstyle\sum_x W(s,x,y,\widehat{s}) - Q_{\widehat{S}|Y}(\widehat{s}|y) &= 0 && \forall s, \widehat{s}, y \\
\textstyle\sum_{\widehat{s}} W(s,x,y,\widehat{s}) - Q_{X|S}(x|s) &= 0 && \forall x, s, y \\
Q_{X|S}(x|s) + Q_{\widehat{S}|Y}(\widehat{s}|y) - W(s,x,y,\widehat{s}) &\le 1 && \forall s, x, y, \widehat{s} \\
Q_{X|S}(x|s) &\ge 0 && \forall s, x \\
Q_{\widehat{S}|Y}(\widehat{s}|y) &\ge 0 && \forall \widehat{s}, y \\
W(s,x,y,\widehat{s}) &\ge 0 && \forall s, x, y, \widehat{s}
\end{aligned}$$

Linear programming relaxation

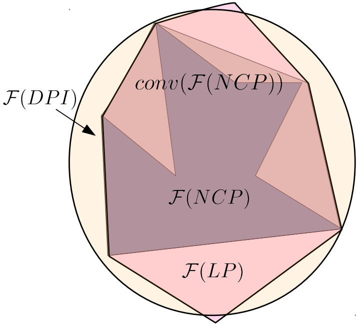
Lemma

All extreme points *of the original problem are extreme points of the LP relaxation.*

- New variables W lift the problem to a higher dimensional space. There it is easier to identify implied inequalities
- Related to techniques in combinatorial optimization – "lift and project", "reformulation-linearization", "method of moments", "McCormick Relaxations", etc.
- Our relaxation *equals the convex hull* in some extreme cases, (e.g., $|\mathcal{S}| = 1$). Exact characterization of the convex hull remains open
- Extendable to networked settings; similar ideas can be applied.



Power of the LP relaxation



1. One can **systematically obtain** lower bounds

Nonclassical problem	$\geq$	**value of LP relaxation**	$=$	**value of dual LP**	$\geq$	**value of any dual feasible point**

2. Can be **intersected** with other relaxation to improve both relaxations
3. Gives new and wider class of **inverse optimal cost functions**

Lossy source-channel coding

- Binary uniform source: $S = (S_1, \cdots S_n)$, where $S_i = \mathrm{Ber}(\frac{1}{2})$, i.i.d.
- Binary symmetric channel: $X = (X_1, \cdots, X_n)$, $Y = (Y_1, \cdots, Y_n)$, where $X_i, Y_i \in \{0, 1\}$.

$$P_{Y_i|X_i}(0|1) = P_{Y_i|X_i}(0|1) = \epsilon < \frac{1}{2}, \qquad P_{Y|X}(y|x) = \Pi_{i=1}^{n} P_{Y_i|X_i}(y_i|x_i)$$

- Cost :

$$\kappa(s, x, y, \widehat{s}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{s_i \neq \widehat{s}_i} \qquad \Longrightarrow \mathbb{E}[\kappa] = \frac{1}{n}\sum_{i=1}^{n} P(S_i \neq \widehat{S}_i)$$

Dual
$$\begin{aligned}
\underset{\gamma^A,\gamma^B,\lambda^A,\lambda^B,\mu}{\text{maximize}} \quad & \sum_s \gamma^A(s) + \sum_y \gamma^B(y) - \sum_z \mu(z) \\
\text{subject to} \quad & \gamma^A(s) - \textstyle\sum_y \lambda^B(x, s, y) - \sum_{\widehat{s}, y} \mu(z) \leq 0 & \forall x, \\
& \gamma^B(y) - \textstyle\sum_s \lambda^A(s, \widehat{s}, y) - \sum_{x, s} \mu(z) \leq 0 & \forall \widehat{s}, \\
& \lambda^A(s, \widehat{s}, y) + \lambda^B(x, s, y) + \mu(z) \leq \tfrac{1}{2^n}\epsilon^{|x-y|}(1-\epsilon)^{n-|x-y|}\tfrac{|s-\widehat{s}|}{n} & \forall z \\
& \mu(z) \geq 0 & \forall z
\end{aligned}$$

$$z := (s, x, y, \widehat{s})$$

$$|s| := \text{Hamming weight of } s$$

$$|x - y| := \text{Hamming distance between } x, y$$

Lossy source-channel coding

Theorem (Tightness for uniform source)

Let $0 < \epsilon < \frac{1}{2}$. The LP relaxation is **tight** *with cost $= \epsilon$ for all n. Moreover, there exists a dual* **optimal** *solution that takes the following form:*

$$\gamma^A(s) \equiv const, \quad \gamma^B(y) \equiv const, \quad \mu(x, s, \widehat{s}, y) \equiv 0$$
$$\lambda^A(s, \widehat{s}, y) \equiv \lambda^A(|s - \widehat{s}|), \quad \lambda^B(x, s, y) \equiv \lambda^B(|x - y|)$$

Proof.

It is easy to show that cost ϵ is achievable in the original problem. Moreover, the following point is feasible:

$$\lambda^A(0) = c \qquad \gamma^A = \sum_{k=0}^{n} {}^nC_k \lambda^B(k), \quad \gamma^B = \sum_{k=0}^{n} {}^nC_k \lambda^A(k)$$

$$\lambda^A(|s - \widehat{s}| = a) = c + \sum_{k=1}^{a} \epsilon^k (1-\epsilon)^{(n-k)} \frac{1}{n \mid S \mid} \qquad \forall a = 1, 2 \ldots n$$

$$\lambda^B(|x - y| = b) = \frac{b}{n \mid S \mid} \epsilon^b (1-\epsilon)^{(n-b)} - \lambda^A(|s - \widehat{s}| = b) \quad \forall b = 0, 1, 2, \cdots n$$

where c is any scalar. The dual cost is, $\mid S \mid \gamma^A + \mid Y \mid \gamma^B = \epsilon$. ∎

Lossy source-channel coding (binary non-uniform source)

More generally ...

- Source (possibly **non-uniform**): $S = (S_1, \cdots S_n)$, $S_i = \mathrm{Ber}(p)$ i.i.d., $p \in [0, 1]$.

Theorem (A general lower bound)

There exists a dual feasible point of the form:

$$\gamma^A(s) \equiv \gamma^A(\mathbf{s}), \quad \gamma^B(y) \equiv const, \quad \mu(x, s, \widehat{s}, y) \equiv 0,$$

$$\lambda^A(s, \widehat{s}, y) \equiv g_{\mathbf{s}}(|s - \widehat{s}|), \quad \lambda^B(x, s, y) = -g_{\mathbf{s}}(|x - y|) + |x - y| g_{\mathbf{s}}'(|x - y|).$$

where $\mathbf{s} \equiv |s|$ *and* $g_{\mathbf{s}} : \mathbb{R} \to \mathbb{R}$ *is* **any concave differentiable function** *satisfying*

$$g_{\mathbf{s}}'(x) \leq \frac{p^{\mathbf{s}}(1 - p)^{n - \mathbf{s}}}{n} \epsilon^x (1 - \epsilon)^{n - x}.$$

We have the following lower bound,

$$\frac{1}{n} \sum_{i=1}^{n} P(S_i \neq \widehat{S}_i) \geq$$

$$\sup_g \left[\sum_{\mathbf{s}=0}^{n} {}^nC_{\mathbf{s}} \sum_{k=0}^{n} {}^nC_k \left[-g_{\mathbf{s}}(k) + k g_{\mathbf{s}}'(k) \right] + 2^n \left[\min_{v \in \{0,1,\cdots n\}} \sum_{i=0}^{n-v} \sum_{j=0}^{v} g_{i-j+v}(i+j)^{n-v} C_i {}^v C_j \right] \right]$$

Channel coding

- Source: $\mathcal{S} = \widehat{\mathcal{S}} \subset \mathbb{N}$, $|\mathcal{S}| < \infty$, $S \sim \mathrm{Unif}(\mathcal{S})$
- Binary symmetric channel: $X = (X_1, \cdots, X_n)$, $Y = (Y_1, \cdots, Y_n)$, where $X_i, Y_i \in \{0, 1\}$.

$$P_{Y_i|X_i}(0|1) = P_{Y_i|X_i}(0|1) = \epsilon < \frac{1}{2}, \qquad P_{Y|X}(y|x) = \Pi_{i=1}^{n} P_{Y_i|X_i}(y_i|x_i)$$

- Cost:

$$\kappa(s, x, y, \widehat{s}) = \mathbb{I}_{s \neq \widehat{s}} \qquad \Longrightarrow \qquad \mathbb{E}[\kappa] = \mathbb{P}(S \neq \widehat{S}) = P(\text{“error''})$$

Theorem (“Strong converse”)

There exists a dual feasible point of the form,

$$\mu \equiv 0, \quad \lambda^A(s, \widehat{s}, y) \equiv \lambda^A(|s - \widehat{s}|), \quad \lambda^B(x, s, y) \equiv \lambda^B(|x - y|)$$
$$\gamma^A(s) \equiv const, \quad \gamma^B(y) \equiv const.$$

Let $|\mathcal{S}| = 2^{nR}$*. If*

$$R > 1 + \log_2(1 - \epsilon),$$

$P(S \neq \widehat{S}) \xrightarrow{n} 1$ **exponentially**, *i.e., there exists* $c \in (0, 1)$ *such that* $P(S \neq \widehat{S}) \geq 1 - c^n$.

Thank You!
Questions, comments?

References

R. Bansal and T. Başar.
Stochastic teams with nonclassical information revisited: When is an affine law optimal?
Automatic Control, IEEE Transactions on, 32(6):554–559, jun 1987.

Emmanuel J Candes and Terence Tao.
Near-optimal signal recovery from random projections: Universal encoding strategies?
Information Theory, IEEE Transactions on, 52(12):5406–5425, 2006.

Sharu Theresa Jose and Ankur A. Kulkarni.
A linear programming relaxation for stochastic control problems with non-classical information patterns.
In *IEEE Conference on Decision and Control*, pages 5743–5748, 2015.

Sharu Theresa Jose and Ankur A. Kulkarni.
Convex-analytic converses for finite-blocklength joint-source channel coding.
to be submitted to the IEEE Transactions on Information Theory, 2016.

A. A. Kulkarni and T. P. Coleman.
An optimizer's approach to stochastic control problems with nonclassical information structures.
In *Proceedings of the IEEE Conference on Decision and Control*, 2012.

A. A. Kulkarni and T. P. Coleman.
An optimizer's approach to stochastic control problems with nonclassical information structures.
IEEE Transactions on Automatic Control, 60(4):937–949, 2015.

Bharat Prabhakar and Ankur A. Kulkarni.
Dimensionality reduction of affine variational inequalities using random projections.
under review with Mathematical Programming. arXiv: http://arxiv.org/abs/1408.4551, 2014.

Bharat Prabhakar and Ankur A. Kulkarni.
Dimensionality reduction of affine variational inequalities using random projections.
In *Proceedings of the Allerton Conference on Communication, Control and Computing*, 2014.

Bharat Prabhakar and Ankur A. Kulkarni.
Dimensionality reduction and sparsification of linear complementarity problems and affine variational using random projections.