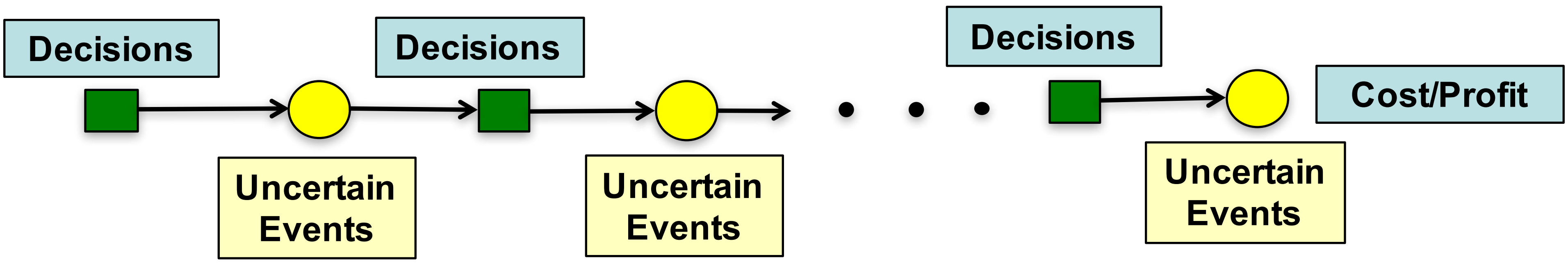
On the Power of Affine Policies in Dynamic Optimization

Vineet Goyal
IEOR, Columbia University

(Joint work with Omar El Housni (Columbia))

BIRS Workshop, March 2018

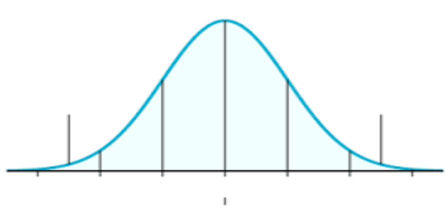
Dynamic Optimization

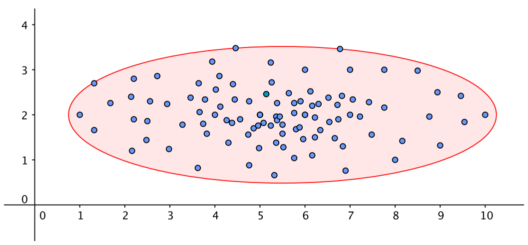


Hybrid Model: Distributionally Robust Optimization

Stochastic optimization

Robust optimization





Computing optimal adjustable policy is intractable

Policy Approximations

- **Static Policies**
  - Single solution feasible for all scenarios
  - Highly tractable but can be very conservative

- **Affine Policy (or Linear Decision Rules)**
  - Recourse solution is an affine function of past uncertainties
  - Tractable and good empirical performance
  - Worst case performance can be bad

- **More general policies**
  - Piecewise static policies
  - Piecewise affine policies
  - Improved performance but significantly more difficult to compute

Policy Approximations

- **Static Policies**
  - Single solution feasible for all scenarios
  - Highly tractable but can be very conservative

- **Affine Policy (or Linear Decision Rules)**
  - Recourse solution is an affine function of past uncertainties
  - Tractable and good empirical performance
  - Worst case performance can be bad

- **More general policies**
  - Piecewise static policies
  - Piecewise affine policies
  - Improved performance but significantly more difficult to compute

Performance of Affine Policies

Provably Optimal for a small class of problems

Worst-case bound $\Theta(\sqrt{\dim(U)})$

How to bridge the gap?

Observed empirical performance is near-optimal

This Talk: We provide a theoretical justification of the contrast between the observed empirical performance and worst-case performance of affine policies

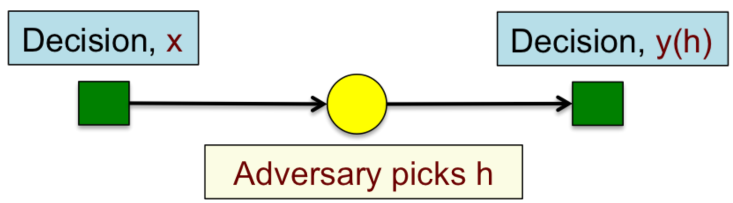
Affine Policies: Empirical Performance

- **Synthetic Data**
  - Randomly generated problem instances
- **Commonly used Uncertainty Sets**
  - Budget of uncertainty sets
  - Intersection of budget of uncertainty sets

This Talk: Analyze Performance of affine for randomly generated instances and for budget of uncertainty sets

Two-stage Adjustable Robust problem

$$z_{\mathsf{AR}} = \min_{\boldsymbol{x}} \boldsymbol{c}^T \boldsymbol{x} + \max_{\boldsymbol{h} \in \mathcal{U}} \min_{\boldsymbol{y}(\boldsymbol{h})} \boldsymbol{d}^T \boldsymbol{y}(\boldsymbol{h})$$
$$\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{h}) \geq \boldsymbol{h}$$
$$\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{h}) \in \mathbb{R}^n_+$$



- Many applications
  - facility location, capacity planning, network design
- computationally intractable in general
- Even approximating LP within an factor of O(log n/log log n) is NP-hard [Feige et al.'07]

Affine Policy approximation

$$\boldsymbol{y}(\boldsymbol{h}) = \boldsymbol{P}\boldsymbol{h} + \boldsymbol{q}$$

Second-stage decision is an affine function of the uncertainty

Affine approximation

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{P},\boldsymbol{q}} \ & \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{h}\in\mathcal{U}} \boldsymbol{d}^T\left(\boldsymbol{P}\boldsymbol{h} + \boldsymbol{q}\right) \\ & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\left(\boldsymbol{P}\boldsymbol{h} + \boldsymbol{q}\right) \ \geq \ \boldsymbol{h} \\ & \boldsymbol{P}\boldsymbol{h} + \boldsymbol{q} \ \geq \ \boldsymbol{0}, \quad \boldsymbol{x} \in \mathbb{R}^n_+ \end{aligned}$$

- Introduced by Ben-Tal et al. (2004)
- Can be computed efficiently
- Optimal for simplex uncertainty sets and very good empirical performance more generally
- Worst case bound is $O(\sqrt{m})$ (Bertsimas and G (2011))
  - Improved bounds for some special uncertainty sets (Bertsimas and Bidkhori (2015))

Random Instances: Performance of Affine Policies

Two-stage Adjustable Problem

$$z_{\mathsf{AR}} = \min_{\boldsymbol{x}} \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{h}\in\mathcal{U}} \min_{\boldsymbol{y}(\boldsymbol{h})} \boldsymbol{d}^T\boldsymbol{y}(\boldsymbol{h})$$

$$\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{h}) \geq \boldsymbol{h}$$

$$\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{h}) \in \mathbb{R}^n_+$$

Affine approximation

$$\min_{\boldsymbol{x},\boldsymbol{P},\boldsymbol{q}} \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{h}\in\mathcal{U}} \boldsymbol{d}^T(\boldsymbol{P}\boldsymbol{h} + \boldsymbol{q})$$

$$\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}(\boldsymbol{P}\boldsymbol{h} + \boldsymbol{q}) \geq \boldsymbol{h}$$

$$\boldsymbol{P}\boldsymbol{h} + \boldsymbol{q} \geq \boldsymbol{0}, \quad \boldsymbol{x} \in \mathbb{R}^n_+$$

Theorem. *Suppose coefficients B_{ij} are i.i.d. according to bounded distribution or with sub-gaussian tails, then affine policy is "near optimal" with high probability for any c, A and polyhedral uncertainty set U*

Random instances with i.i.d. bounded distributions

Suppose B_{ij} are i.i.d. according to a bounded distribution with support in $[0,b]$ and $\mathbb{E}(B_{ij})=\mu$

Theorem. For n sufficiently large compared to $\log m$, with probability at least $1-\frac{1}{m}$, we have

$$z_{AR} \leq z_{Aff} \leq \frac{b}{\mu(1-\epsilon)} \; z_{AR}$$

Examples:

- **B_{ij} are i.i.d. Uniform [0,1]:**
  Affine policy gives a 2-approximation to the two-stage adjustable problem

- **B_{ij} are i.i.d. Bernoulli(p):**
  Affine policy gives a $\frac{1}{p}$-approximation to the two-stage adjustable problem.

Random instances with i.i.d. unbounded distributions

Suppose B_{ij} are i.i.d. according to absolute value of a standard Gaussian distribution

Theorem. For n sufficiently large compared to $\log m$, with probability at least $1 - \frac{1}{m}$, we have

$$z_{AR} \leq z_{Aff} \leq \kappa . z_{AR}$$

where $\kappa = O\left(\sqrt{\log m + \log n}\right)$

- Result extends to distributions with sub-gaussian tails

Proof (Sketch)

Based on duality in constraints and uncertainty set (Bertsimas and de Ruiter (2016))

Primal two-stage problem

$$z_{\text{AR}} = \min_{\boldsymbol{x}} \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{h}\in\mathcal{U}} \min_{\boldsymbol{y}(\boldsymbol{h})} \boldsymbol{d}^T\boldsymbol{y}(\boldsymbol{h})$$
$$\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{h}) \geq \boldsymbol{h}$$
$$\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{h}) \in \mathbb{R}^n_+$$

Primal uncertainty set

$$\mathcal{U} = \{\boldsymbol{h} \in \mathbb{R}^m_+ \mid \boldsymbol{R}\boldsymbol{h} \leq \boldsymbol{r}\}$$

=

Dual two-stage problem

$$\min_{\boldsymbol{x}} \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{w}\in\mathcal{W}} \min_{\boldsymbol{\lambda}(\boldsymbol{w})} -(\boldsymbol{A}\boldsymbol{x})^T\boldsymbol{w} + \boldsymbol{r}^T\boldsymbol{\lambda}(\boldsymbol{w})$$
$$\boldsymbol{R}^T\boldsymbol{\lambda}(\boldsymbol{w}) \geq \boldsymbol{w}$$
$$\boldsymbol{\lambda}(\boldsymbol{w}) \in \mathbb{R}^L_+, \ \boldsymbol{x} \in \mathbb{R}^n_+$$

Dual uncertainty set

$$\mathcal{W} = \{\boldsymbol{w} \in \mathbb{R}^m_+ \mid \boldsymbol{B}^T\boldsymbol{w} \leq \boldsymbol{d}\}$$

Theorem [Bertsimas and De ruiter 2016] : Affine approximation of the primal and dual are equivalent

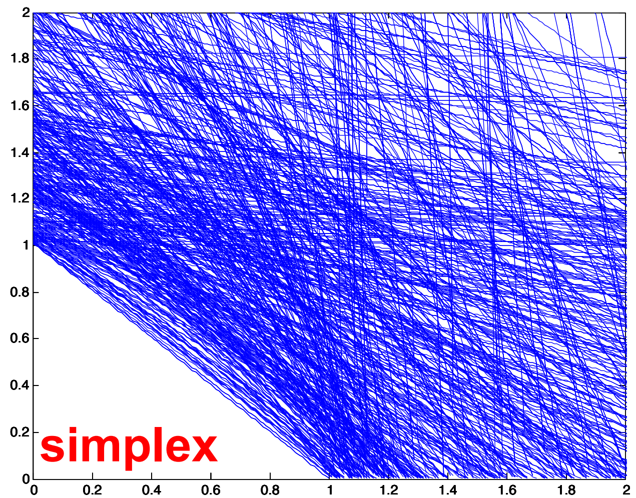
→ We get a new two-stage adjustable problem where uncertainty set depends on the random matrix $\boldsymbol{B}$

Proof (Sketch)

$$\mathcal{W} = \{\boldsymbol{w} \in \mathbb{R}^m_+ \mid \boldsymbol{B}^T \boldsymbol{w} \leq \boldsymbol{d}\}$$

We show with high probability that $\mathcal{W}$ can be approximated by a simplex when B_{ij} are i.i.d.

Example:



B_{ij} i.i.d. Uniform [0,1]

Near-optimality of affine policies follows from the optimality for simplex uncertainty sets

Numerical Performance

Comparison of affine and adjustable policy in terms of performance and running times

B_{ij} i.i.d. Uniform [0,1]

m	r_{avg}	r_{max}	$T_{\text{AR}}(s)$	$T_{\text{Aff}}(s)$
10	1.01	1.03	10.55	0.01
20	1.02	1.04	110.57	0.23
30	1.01	1.02	761.21	1.29
50	**	**	**	14.92

(a) Uniform

B_{ij} i.i.d. Folded Normal

m	r_{avg}	r_{max}	$T_{\text{AR}}(s)$	$T_{\text{Aff}}(s)$
10	1.00	1.03	12.95	0.01
20	1.01	1.03	217.08	0.39
30	1.01	1.03	594.15	1.15
50	**	**	**	13.87

(b) Folded Normal

No Smoothed Analysis: Family of Bad Instances

Family of bad instances

$n = m, \quad \boldsymbol{A} = \boldsymbol{0}, \quad \boldsymbol{c} = \boldsymbol{0}, \quad \boldsymbol{d} = \boldsymbol{e}$

$\mathcal{U} = \text{conv}\left(\boldsymbol{0}, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_m\right)$ where $\boldsymbol{\nu}_i = \frac{1}{\sqrt{m}}(\boldsymbol{e} - \boldsymbol{e}_i)\ \forall i \in [m]$.

$\tilde{B}_{ij} = \begin{cases} 1 & \text{if } i = j \\ \frac{1}{\sqrt{m}} \cdot \tilde{u}_{ij} & \text{if } i \neq j \end{cases}$ where for all $i \neq j$, $\tilde{u}_{ij}$ are i.i.d. uniform$[0,1]$.

Coefficients are not i.i.d. !!!

Theorem. For the above instance, we have with probability at least $1 - \frac{1}{m}$,

$$z_{Aff} = \Omega(\sqrt{m}).\, z_{AR}$$

Performance of Affine Policies

- Real world instances are Not random
  - Affine policies exhibit good empirical performance more generally

- **Commonly used Uncertainty Set**

Budget of Uncertainty Set:

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^{m} w_i h_i \le \Gamma \right\}$$

- Very commonly used class of uncertainty sets
- More general: intersection of budget of uncertainty sets
- Captures confidence interval sets and CLT based sets

Hardness (Feige et al. 2007): Adjustable problem is hard to approximate within a factor $\Omega\left(\frac{\log n}{\log \log n}\right)$ for budget of uncertainty sets.

Performance of Affine Policies

Budget of Uncertainty Set:

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^{m} w_i h_i \leq \Gamma \right\}$$

Theorem. *Affine policy gives $O(log\, n)$-approximation for budget of uncertainty sets*

Optimal approximation: nearly matches the hardness bound

Intersection of Budget of Uncertainty Sets

- **Partition Matroid** (Intersection of Budget of disjoint subsets)
  - Generalization of budget of uncertainty
  - $I_1, I_2, \ldots, I_L$ is a partition of [m].

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i \in I_\ell} h_i \le k_\ell \;\; \forall \ell = 1, \ldots, L \right\}$$

Theorem. *Affine policy gives $O(\log^2 n)$-approximation for partition matroid uncertainty sets*

Intersection of Budget of Uncertainty Sets

Theorem. *For U given by intersection of L budget constraints, affine policy gives:*

- $O(log\, n\, log\, L)$*-approximation if U is permutation invariant*
- $O(L\, log\, n)$*-approximation in general.*

- Example of permutation invariant budgeted set: **CLT based set**

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i \in \mathcal{S}} h_i \leq \gamma \;\; \forall \mathcal{S} \subseteq [m] \text{ with } |\mathcal{S}| = k \right\}$$

Special Constraint Matrix: $\boldsymbol{B}$ Totally-Unimodular

Theorem. If the second-stage constraint matrix is totally unimodular, affine policy gives a 5-approximation for budget of uncertainty sets.

- Many applications where $\boldsymbol{B}$ is TU
  - facility location
  - transportation problems
  - supply chain network design

- The bounds also extend to the case of intersection of L budget sets
  - $O(\log L)$ for permutation invariant sets
  - $O(L)$ for general intersection of budgeted sets

Proof (Sketch for budget of uncertainty set)

Budget of Uncertainty Set

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^{m} h_i \leq k \right\}$$

- Show existence of a good affine solution.
- Exploit the instance constraints: A, B and costs: c, d unlike analysis in prior work

Proof (Sketch)

Budget of Uncertainty Set

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^{m} h_i \leq k \right\}$$

Step 1: Pruning Inexpensive Scenarios

$$\theta_i = \min_{\boldsymbol{y}} \left\{ \boldsymbol{d}^T \boldsymbol{y} \;\middle|\; \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{e}_i, \quad \boldsymbol{y} \geq \boldsymbol{0} \right\}$$

$y^*(e_i)$ **: Optimal solution**

$$\mathcal{I}_1 = \left\{ i \in [m] \;\middle|\; \theta_i \leq O(\log n) \cdot \frac{\mathbf{OPT}}{k} \right\}$$

Cover all components in I_1 in second stage by a linear solution

$$y(h) = \sum_{i \in \mathcal{I}_1} y^*(e_i) \cdot h_i$$

Cost increases by a factor log n

Proof (Sketch): Remaining components

Step 2 (Remaining Components) $\mathcal{I}_2 = [m] \setminus \mathcal{I}_1$

- cover remaining components using a static solution

$$\hat{x} \in \textbf{argmin} \left\{ d^T x \;\middle|\; Bx \geq \sum_{i \in \mathcal{I}_2} e_i, \;\; x \geq 0 \right\}$$

What about the cost of $\hat{x}$?

Lemma: Cost of $\hat{x}$ is at most O(OPT).

- Each remaining component is more than (log n OPT)/K
- Total cost of any subset of size K is at most OPT
- Using these two properties we show the existence of a good solution
  - Adapt arguments from Gupta et al. (2011))

Faster algorithm for Approximate affine policies

- Based on insights from the proof of performance bounds

$$\theta_i = \min_{\boldsymbol{y}} \left\{ \boldsymbol{d}^T \boldsymbol{y} \;\middle|\; \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{e}_i, \;\; \boldsymbol{y} \geq \boldsymbol{0} \right\}$$

Suppose $\boldsymbol{\theta_1 \geq \theta_2 \geq \ldots \geq \theta_m}$

- Try the following m affine solutions

- For j =1…m
  - Cover $e_1, \ldots, e_j$ with a static first stage solution
  - Affine solution:

$$\boldsymbol{y}(\boldsymbol{h}) = \sum_{i=j+1}^{m} \boldsymbol{y}^*(\boldsymbol{e}_i) \cdot \boldsymbol{h}_i$$

- Return the solution with minimum cost

Numerical Performance of Faster algorithm

m	$T_{\mathsf{aff}}(s)$	$T_{\mathsf{Alg}}(s)$	Alg/Aff
10	0.009	0.004	1.146
20	0.176	0.011	1.106
30	0.587	0.024	1.143
40	2.395	0.039	1.145
50	9.718	0.063	1.097
60	17.40	0.087	1.155
70	52.36	0.118	1.101
80	108.8	0.155	1.128
90	188.7	0.205	1.133
100	270.7	0.247	1.146

Conclusions

- Affine policies are **Near-optimal** for random instances generated from a large class of distribution
- Provide **Optimal approximation** for budget of uncertainty sets that are widely used in practice
- Faster algorithm to compute near-optimal affine policies
- **Extend insights to more general policies**

Thank You.

References

[1] O. El Housni and V. Goyal. Beyond Worst-case: A Probabilistic Analysis of Affine Policies in Dynamic Optimization. *In NIPS (2017)*

[2] O. El Housni and V. Goyal. Optimal Approximation using Affine Policies for Budget of Uncertainty Sets. *In preparation*

Questions?