Voyage incertain

Découvrir l’optimisation stochastique et robuste

Pourquoi de l’incertitude ?

- Un modèle d’optimisation dépend souvent de paramètres **incertains**
  - Demande d’un client
  - Panne dans un réseau : lien ou routeur

- Quelles conséquences en pratique ?
  - Des décisions deviennent suboptimales
  - Des décisions deviennent irréalisables

Pourquoi de l’incertitude ?

- Un autre exemple : la bibliothèque de problèmes NETLIB
  - 93 problèmes linéaires réels
  - Changer des coefficients de 0,1 % peut rendre des solutions irréalisables de 450 %

- Ben-Tal, A., & Nemirovski, A. (2000). Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, *88*(3), 411–424.
- http://www.netlib.org/lp/data/index.html

Optimisation stochastique et robuste

Deux approches principales pour gérer l'incertitude :

- Variables aléatoires : **optimisation stochastique**
  - Approche historique : date de 1955 (simplexe : 1947)

    Dantzig, G. B. (2011). Linear Programming Under Uncertainty. In G. Infanger (Ed.), *Stochastic Programming* (Vol. 150, pp. 1–11). Springer New York.
- Ensemble d'incertitude : **optimisation robuste**
  - Inspiré de la théorie des jeux : la meilleure solution est celle qui minimise le risque maximal (principe du maximin)

    Wald, A. (1945). Statistical decision functions which minimize the maximum risk. *Annals of Mathematics*, 265–280.

Optimisation stochastique et robuste

Application dans le routage

- **Approche stochastique** :
  - Déterminer une densité de probabilité depuis les données pour les demandes
  - Minimiser la congestion **moyenne** dans le réseau pour ces demandes
- **Approche robuste** :
  - Déterminer toutes les combinaisons de demandes possibles
  - Minimiser la congestion **maximale** dans le réseau

Table des matières

Optimisation stochastique

Optimisation stochastique

- Idée principale : les paramètres incertains sont donnés par une densité de probabilité

- L’objectif est une **mesure de risque**
  - Le plus souvent une espérance
  - Si les variations peuvent être importantes : ajouter un terme de variance, par exemple

Formulation

- L'espérance est définie comme une intégrale
  - Difficile à calculer !
- En pratique : la densité de probabilité est discrétisée par un ensemble de scénarios

$$\mathbb{E}_{\boldsymbol{w}}\{f(x, y, w)\} \approx \sum_{s} p_s \, f(x, y_s, w_s)$$

Formulation

Deux types de variables :

- Les décisions à prendre **avant** que l'incertitude soit révélée
  - Quantité de journaux à commander la veille

$$\Rightarrow x$$

- Les « décisions » à prendre **après**
  - Quantité de journaux vendue en fin de journée

$$\Rightarrow y_s$$

Exemple : placement d’infrastructure

- Le problème : servir des clients depuis une série d’emplacements potentiels (CFLP)
  - Utiliser un emplacement a un coût
  - Servir un client a un coût qui dépend de l’emplacement choisi
  - La demande des clients n’est pas exactement connue

- Par exemple, positionner des caches dans un réseau

Exemple : placement d’infrastructure

La formulation classique :

- x_i indique si l’emplacement *i* est utilisé (binaire)
- $y_{i,j}$ indique la demande du client *j* servie par *i*

$$\begin{array}{ll} \text{min} & \displaystyle\sum_i f_i\, x_i + \sum_i \sum_j c_{i,j}\, y_{i,j} \\ \text{telle que} & \displaystyle\sum_j y_{i,j} \leq \text{capa}_i\, x_i, \text{pour tout serveur } i \\ & \displaystyle\sum_i y_{i,j} \geq \text{dem}_j, \text{pour tout client } j \end{array}$$

Exemple : placement d'infrastructure

Une formulation stochastique :

- x_i indique si l'emplacement *i* est utilisé (binaire)
- $y_{i,j}^s$ indique la demande du client *j* servie par *i* **pour le scénario s**

$$\min \quad \sum_i f_i\, x_i + \sum_s p_s \sum_i \sum_j c_{i,j}\, y_{i,j}^s$$

$$\text{telle que} \quad \sum_j y_{i,j}^s \leq \text{capa}_i\, x_i, \text{pour tout serveur } i \text{ et tout scénario } s$$

$$\sum_i y_{i,j}^s \geq \text{dem}_j^s, \text{pour tout client } j \text{ et tout scénario } s$$

Résolution

- La taille d'un problème stochastique augmente vite
  - Linéairement avec le nombre de scénarios
  - Avec des milliers ou des millions de scénarios…
- Plusieurs techniques de résolution :
  - Équivalent déterministe, comme précédemment
  - Décomposition : souvent Benders
  - Heuristiques : par exemple, recouvrement progressif

Résolution : décomposition de Benders

- Un programme stochastique a une structure très forte :
  - Les scénarios sont indépendants les uns des autres
  - Tous les scenarios dépendent des mêmes variables
- La décomposition de Benders isole ces groupes
  - Collaboration entre chaque scénario et un « maître »

Infanger, G. (1992). *Planning under uncertainty solving large-scale stochastic linear programs*.
Murphy, J. (2013). Benders, Nested Benders and Stochastic Programming: An Intuitive Introduction. *ArXiv E-Prints*, 57. Retrieved from http://arxiv.org/abs/1312.3158

Autre nom : « méthode en L »

Résolution : décomposition de Benders

- Schématiquement, pour le placement de caches :
  - Le programme maître :
$$\min \sum_i o_i \, x_i + q(y)$$
  - Les programmes par scénario :
$$\min \sum_i \sum_j c_{i,j} \, y_{i,j}$$

Résolution : décomposition de Benders

L’algorithme fonctionne comme suit :

- Résoudre le problème maître
  $$\Rightarrow x$$
- Résoudre les problèmes pour chacun des scénarios (indépendamment ou non)
  $$\Rightarrow y^s \text{ en fonction de } x$$
- Ajouter des contraintes au problème maître selon les solutions à chacun des scénarios
  $$\Rightarrow \text{informations sur } q(y)$$
- Itérer

Résolution : recouvrement progressif

Autre idée pour exploiter la structure :

- Optimiser les scénarios indépendamment les uns des autres

$$\Rightarrow x^s, y^s$$

- Ajouter une pénalité pour rapprocher les solutions

$$\Rightarrow \bar{x} = \sum_s p_s x^s$$

$$\Rightarrow \min(c^s)^T x^s + (d^s)^T y^s + ||x^s - \bar{x}||$$

Watson, J.-P., & Woodruff, D. L. (2011). Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems. *Computational Management Science*, *8*(4), 355–370. http://doi.org/10.1007/s10287-010-0125-4

Optimisation robuste

Optimisation robuste

- Idée principale : optimiser **dans le pire cas**
  - Les paramètres sont donnés par un ensemble d'incertitude

- Formellement :

$$\min_{x \in \mathcal{X}} \max_{w \in \mathcal{W}} f(x, w)$$

Ensembles d'incertitude courants

Bon nombre de modèles d'incertitude ont été proposés, voici les trois principaux :

- Boîte (Soyster) :

$$w_i \in [w_{\min}, w_{\max}]$$

- Ellipsoïde (Ben-Tal et Nemirovski) :

$$\sum_i \widehat{w}_i^2 \leq \Omega$$

- Nombre de changements (Bertsimas et Sim)

$$(\#\widehat{w}_i \neq 0) \leq \Omega$$

Ensembles d’incertitude courants

- Modèle de Soyster :
  - Soyster, A. L. (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, *21*(5), 1154–1157.
- Modèle de Ben-Tal et Nemirovski :
  - Ben-Tal, A., & Nemirovski, A. (2000). Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, *88*(3), 411–424.
- Modèle de Bertsimas et Sim :
  - Bertsimas, D., & Sim, M. (2004). The price of robustness. *Operations Research*, *52*(1), 35–53.

Formulation

- En l’état, la programmation robuste fait appel à un nombre **infini** de contraintes :

$$\min_{x \in \mathcal{X}} u \ \text{ telle que } u \geq f(x, w), \forall w$$

- Deux techniques pour résoudre un problème robuste :
  - Reformulation : risque de perdre la linéarité du problème
  - Plans sécants : solution nettement plus générique

Bertsimas, D., Dunning, I., & Lubin, M. (2014). Reformulations versus cutting planes for robust optimization. *Optimization Online*.

Exemple : placement d'infrastructure

- Avec le modèle de Soyster, chaque demande est dans une boîte : $\mathrm{dem}_j \in \left[\underline{\mathrm{dem}}_j, \overline{\mathrm{dem}}_j\right]$
- Reformulation équivalente :

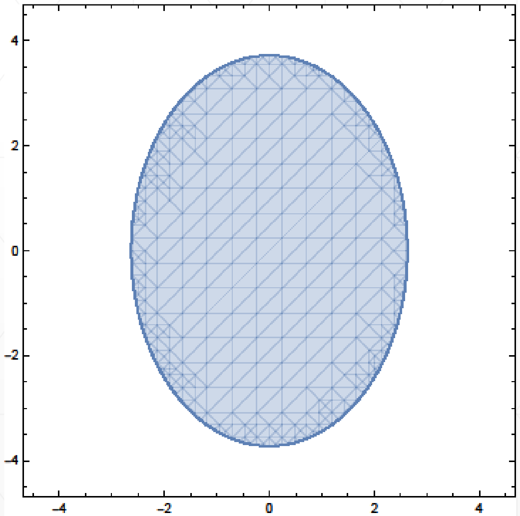
$$\min \quad \sum_i f_i\, x_i + \sum_i \sum_j c_{i,j}\, y_{i,j}$$

$$\text{telle que} \quad \sum_j y_{i,j} \leq \mathrm{capa}_i\, x_i, \text{pour tout serveur j}$$

$$\sum_i y_{i,j} \geq \boxed{\overline{\mathrm{dem}}_j}, \text{pour tout client i}$$

Exemple : placement d'infrastructure

- Avec le modèle de Ben-Tal et Nemirovski, toutes les demandes sont dans une ellipsoïde
- Toutes les demandes ne sont pas au maximum de leur incertitude en même temps !

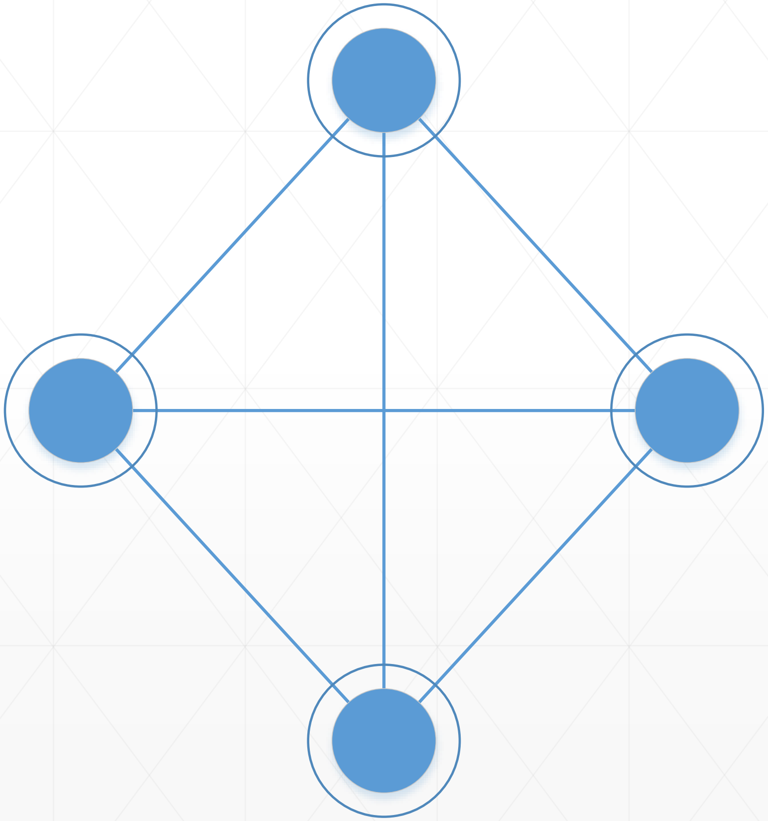


Exemple : routage robuste et VPN

- Problème : dimensionner une infrastructure réseau
  - Le client doit connecter différents sites
  - Il ne peut pas estimer correctement les besoins pour chaque paire de sites
  - Par contre, il connaît assez bien les besoins totaux par site
- Ensemble d'incertitude : **modèle du tuyau flexible**
  - Tous les débits point à point qui respectent les débits totaux par point

Kumar, A., Rastogi, R., Silberschatz, A., & Yener, B. (2002). Algorithms for provisioning virtual private networks in the hose model. *IEEE/ACM Transactions on Networking (TON)*, *10*(4), 565–578.

Exemple : routage robuste et VPN



Résolution : plans sécants

L'algorithme fonctionne comme suit :

- Résoudre le problème initial,
  avec la contrainte $a^T x \leq b$:
  $$\Rightarrow x^*$$
- Résoudre le sous-problème $\max_{\tilde{a} \in \mathcal{A}} \tilde{a}^T x$
  $$\Rightarrow \tilde{a}^*$$
- Si ${\tilde{a}^*}^T x > b$:
  - Ajouter la contrainte ${\tilde{a}^*}^T x \leq b$ et itérer
- Sinon : solution optimale robuste trouvée !

Application aux files d’attente

- Analyse de la performance par optimisation robuste
  - Étude du pire temps d’attente au niveau d’un serveur
  - Les temps entre deux paquets et de traitement d’un paquet sont décrits par un ensemble d’incertitude

Bandi, C., Bertsimas, D., & Youssef, N. (2015). Robust queueing theory. *Operations Research*, *63*(3), 676–700.

Décisions séquentielles

Côté stochastique

Modèle échelonné

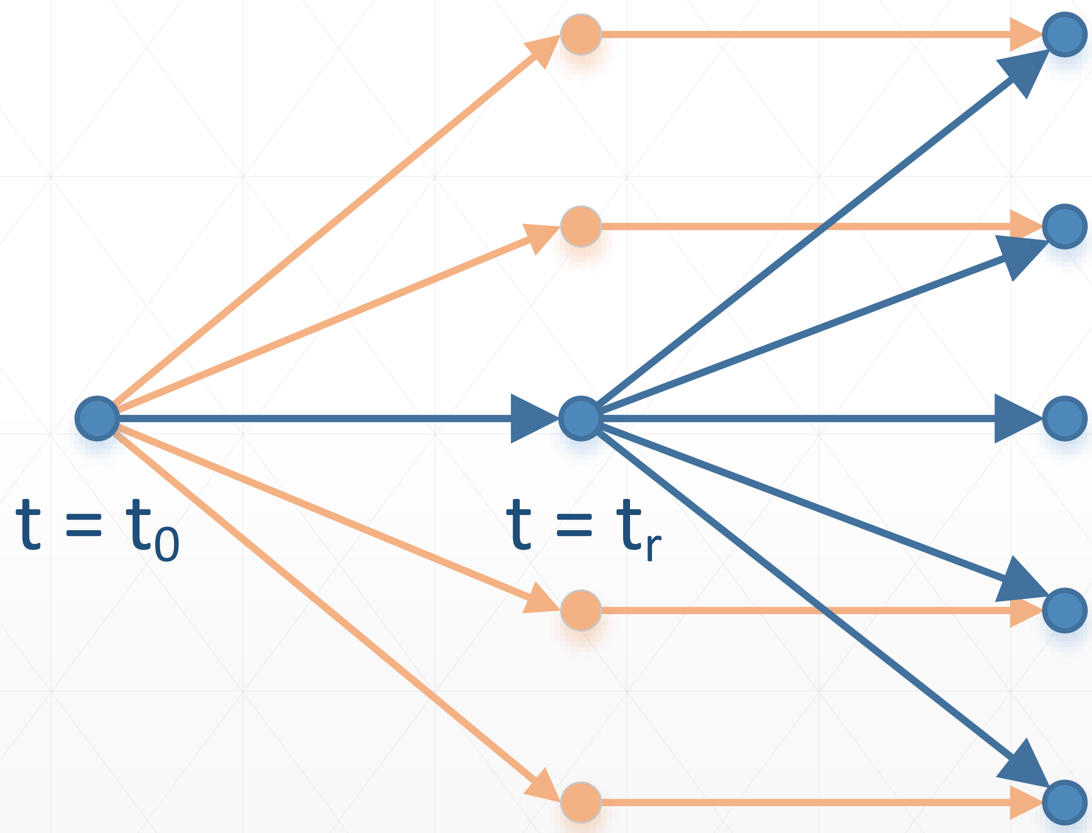
- Toutes les décisions ne doivent pas être prises maintenant
  - Recours futur : prendre des décisions plus tard, *quand une partie de l'incertitude est révélée*
- L'incertitude devient un **arbre de scénarios**
  - Toutes les réalisations possibles de l'incertitude en fonction du passé

Modèle échelonné

Intérêt ?

- Les décisions en t ne doivent pas résister à toutes sortes de scénarios
- Elles doivent juste permettre de s’adapter en temps voulu

Arbre des scénarios



Formulation

- Distinction entre les décisions à prendre maintenant et à chaque pas de temps

$$\min \sum_{s_1} p_{s_1} \Big[f\left(x, y_{s_1}, w_{s_1}\right)$$

Évaluation des décisions **avant** le recours

Évaluation des décisions **après** le recours

Résolution

- Équivalent déterministe
  - Exponentiel en le nombre de pas de temps…
- Décomposition de Benders :
  - Application récursive : on ne considère que le pas de temps suivant pour la décomposition
  - Programmation dynamique (formulation duale : SDDP)
- Recouvrement progressif :
  - Un programme d’optimisation par feuille de l’arbre des scénarios, un terme de pénalité par niveau

Résolution

- Décomposition de Benders :
  - Egging, R. (2013). Benders decomposition for multi-stage stochastic mixed complementarity problems--Applied to a global natural gas market model. *European Journal of Operational Research*, *226*(2), 341–353.
- SDDP :
  - Pereira, M. V. F., & Pinto, L. M. V. G. (1991). Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, *52*(1–3), 359–375.

Décisions séquentielles

Côté robuste

Optimisation ajustable

- Une solution peut parfois être ajustée après la prise de décision
  - Décisions de réassort, par exemple
- Optimisation linéairement ajustable :
  - L'ajustement est une fonction affine
$$y = \alpha\, w + \beta$$
  - L'optimisation porte sur les coefficients α et β

- Autre solution : processus de Markov robuste

Contraintes en probabilité

Retour au stochastique

Contraintes en probabilité

- Qui dit variable aléatoire… dit probabilité !
- D’où : contraintes sur la probabilité d’occurrence de certains événements
  - En fonction des décisions prises

$$P_{\mathcal{W}}(w^T x \geq \alpha) \leq \beta$$

Contraintes en probabilité

- Pas la panacée : contrainte rarement convexe
  - Sauf distributions dont le logarithme est concave
- De plus, exceptionnellement conique
  - Sauf la gaussienne !
- Sinon ? Approximations de la contrainte…

Quid de l'optimisation robuste ?

- Une gaussienne n'est pas toujours un modèle précis de l'incertitude
  - Et si ses paramètres étaient incertains ?
  - D'où un ensemble d'incertitude (optimisation robuste) :
$$\mathcal{X} \sim \mathcal{N}(\mu, \sigma), \qquad \text{avec } (\mu, \sigma) \in \mathcal{U}$$
  - Résolution par plans sécants et linéarisation

Lubin, M., Dvorkin, Y., & Backhaus, S. (2016). A robust approach to chance constrained optimal power flow with renewable generation. *IEEE Transactions on Power Systems*, *31*(5), 3840–3849.

Mot de la fin

Conclusion

- Deux paradigmes pour l’optimisation sous incertitude :
  - Stochastique : variable aléatoire, minimiser l’espérance, contrainte en probabilité
  - Robuste : ensemble d’incertitude, minimiser le pire cas
- Bon nombre d’applications pour les deux
  - Pas d’interdiction de mélanger les deux méthodes !

Des références principales ?

- Optimisation stochastique :
  - Birge, J. R., & Louveaux, F. (2011). *Introduction to Stochastic Programming*. (T. V Mikosch, S. I. Resnick, & S. M. Robinson, Eds.)*Springer Series in Operations Research and Financial Engineering* (Second). Springer Verlag. http://doi.org/10.1007/978-1-4614-0237-4
- Optimisation robuste :
  - Ben-Tal, A., El Ghaoui, L., & Nemirovski, A. (2009). *Robust optimization*. Princeton University Press.

Des questions?