Optimisation non-linéaire globale avec application aux modèles de choix discret

Michaël Thémans Nicolas Zufferey
Michel Bierlaire

Transport and Mobility Laboratory
École Polytechnique Fédérale de Lausanne

Plan

- Motivation: Estimation de modèles de choix discret
  - Fonction-objectif hautement non-linéaire et non-concave
  - Nombreux optima locaux
- Optimisation non-linéaire globale
  - Littérature
  - Approche différente
- Caractéristiques du nouvel algorithme
- Résultats numériques
- Conclusions et perspectives

Motivation

- Modèles de choix discret
  - Utilisés pour analyser et prédire le comportement de choix d'individus
  - Paramètres calibrés sur les choix observés d'individus
- Calibration de modèles de choix discret
  - Estimation du maximum de vraisemblance
  - Problèmes d'optimisation non-linéaire sans contrainte
  - Différentes sources de difficulté pour des modèles avancés

Difficultés dans l'estimation

- La fonction-objectif est hautement non-linéaire et non-concave
- Le problème d'optimisation associé peut être singulier (matrice hessienne non inversible à la solution)
- La fonction-objectif est souvent très coûteuse à évaluer (recours à des outils de simulation)
- La fonction-objectif peut présenter plusieurs (et souvent de nombreux) optima locaux

$\Rightarrow$ L'estimation nécessite des algorithmes d'optimisation spécifiques

Optimisation non-linéaire globale

- Problème peu étudié en économétrie
- Littérature importante en recherche opérationnelle
- Nombreux algorithmes inspirés de l'optimisation discrète
  - Adaption de méta-heuristiques (telles que Recuit Simulé, Tabou, Algorithme Génétique) pour le cas continu
  - Hybridation de méta-heuristiques avec
    - des méthodes sans dérivée (telles que la recherche directe de N-M) comme recherche locale
    - des recherches aléatoires

Nouvelle approche

- Recherche à Voisinage Variable (RVV) pour explorer et diversifier
- 2 éléments essentiels dans la RVV
  - Recherche locale utilisée
  - Définition des voisinages et des voisins
- Utiliser au mieux l'information sur la fonction-objectif et ses dérivées obtenues à moindre coût
- Limiter les évaluations de f en identifiant les zones de recherche prometteuses et celles peu intéressantes

RVV pour le cas continu

- Sélectionner un ensemble de voisinages $\mathcal{N}_k$, $k = 1, \ldots, k_{max}$ et une solution initiale x_*^0 (minimum local)
- $x_c = x_*^0$ et $x_*^{best} = x_*^0$
- $k = 1$
- Tant que $k \leq k_{max}$
  - Générer p voisins de x_c dans $\mathcal{N}_k$
  - Appliquer k_2 itérations de la recherche locale
  - Appliquer la recherche locale au meilleur point obtenu précédemment pour trouver un minimum local x_*^{new}
  - Si $f(x_*^{new}) < f(x_*^{best})$, $x_*^{best} = x_c = x_*^{new}$ et $k = 1$. Sinon $k = k + 1$.
- La solution est x_*^{best}

Recherche locale

- Méthode de région de confiance
- Méthode itérative pour l'optimisation non-linéaire sans contrainte
- Capable de trouver efficacement un minimum local du problème
- Technique de globalisation: convergence à partir de n'importe quel point de départ (pas nécessaire d'être dans le voisinage d'un minimum local)
- Utilisation d'un modèle quadratique de f
  - ∇f approximé par différences finies
  - $\nabla^2 f$ approximé par méthode sécante (quasi-N) SR1

Zones de recherche prometteuses

- Recherche locale performante mais coûteuse
- Éviter d'appliquer la recherche locale inutilement et limiter le nombre d'évaluations de f
- Appliquer des tests à chaque itéré x_k généré par la recherche locale
- Étant donnés $X_* = \{x_*^0, x_*^1, \ldots\}$ et f_{min}
  - $\exists\, i$ tel que $\|x_k - x_*^i\| \leq \varepsilon_1$
  - $\|\nabla f(x_k)\| \leq \varepsilon_2$ et $f(x_k) - f_{min} \geq \varepsilon_3$
  - $f(x_k) > f(x_{k-1}) + \beta \nabla f(x_{k-1})^T s_{k-1}$ et $f(x_k) - f_{min} \geq \varepsilon_3$
- Si un des tests est réussi, on arrête la recherche locale

Voisinages de la RVV

- Le voisinage $\mathcal{N}_k$ est défini par une distance d_k
- Analyse de la structure propre de $\nabla^2 f(x_c) \in \mathbb{R}^{n\times n}$: calcul des n vecteurs propres q_i et des valeurs propres associées c_i (courbure)
- $2n$ directions possibles pour déterminer un voisin
- Idée: privilégier les directions associées à une courbure importante mais...
  - Information locale, autour de x_c
  - Réduire l'importance de la courbure dans le choix des voisins pour des voisinages de taille importante

Voisinages de la RVV

- Probabilité de type logit pour le choix d’une direction:

$$P(q_i) = P(-q_i) = \frac{e^{\lambda \frac{c_i}{d_k}}}{2 \sum_{j=1}^{n} e^{\lambda \frac{c_i}{d_k}}}$$

  où λ est un facteur de poids associé à la courbure

- p directions sont choisies aléatoirement à chaque cycle de la RVV selon le vecteur de probabilité P
- Si q_i est choisi, le voisin associé est donné par:

$$x_c + \alpha * d_k * q_i$$

  où $\alpha \in U([0.75, 1])$

Nouvelle RVV

- Sélectionner un ensemble de voisinages $\mathcal{N}_k$, $k = 1, \ldots, k_{max}$ et une solution initiale x_*^0 (minimum local)
- $x_c = x_*^0$ et $x_*^{best} = x_*^0$
- $k = 1$
- Tant que le critère d'arrêt n'est pas atteint
  - Générer p voisins de x_c dans $\mathcal{N}_k$
  - Appliquer la recherche locale à chaque voisin tant qu'une zone prometteuse est détectée
  - Si toutes les recherches locales sont arrêtées, $k = k + 1$
  - Sinon x_*^{new} est le meilleur minimum local trouvé. Si $f(x_*^{new}) < f(x_*^{best})$, $x_*^{best} = x_c = x_*^{new}$ et $k = 1$. Sinon $k = k + 1$.
- La solution est x_*^{best}

Quelques détails algorithmiques

- Initialisation:
  - Générer m points aléatoirement et appliquer k_2 itérations de la recherche locale
  - Appliquer la recherche locale au meilleur point obtenu
- Critère d'arrêt:
  - $k > k_{max}$ (arrêt après k_{max} cycles RVV infructueux)
  - Temps CPU maximum atteint
  - Nombre d'évaluations de f maximum atteint

Tests numériques préliminaires

- 19 problèmes-test classiques
- 5 algorithmes/compétiteurs
  - RVV proposée
  - Recherche tabou hybridée avec Recherche Directe (DTS, 2006)
  - Recherche par Motifs hybridée avec Recuit Simulé (SAHPS, 2004)
  - Algorithme Génétique hybridé avec Nelder-Mead (CHA, 2003)
  - Recuit Simulé hybridé avec Recherche Directe (DSSA, 2002)
- Mesure de performance: compromis entre
  - Nombre moyen d'évaluations de la fonction-objectif f
  - Pourcentage de succès

Pourcentage de succès

Problem	RVV	CHA (2003)	DSSA (2002)	DTS (2006)	SAHPS (2004)
RC	100	100	100	100	100
ES	100	100	93	82	96
RT	85	100	100		100
SH	80	100	94	92	86
R_2	100	100	100	100	100
Z_2	100	100	100	100	100
DJ	100	100	100	100	100
$H_{3,4}$	100	100	100	100	95
$S_{4,5}$	100	85	81	75	48
$S_{4,7}$	100	85	84	65	57
$S_{4,10}$	100	85	77	52	48
R_5	100	100	100	85	91
Z_5	100	100	100	100	100
$H_{6,4}$	100	100	92	83	72
R_{10}	100	83	100	85	87
Z_{10}	100	100	100	100	100
HM	100		100		
GR	100		90		
CV	100		100		

Nombre moyen d'évaluations

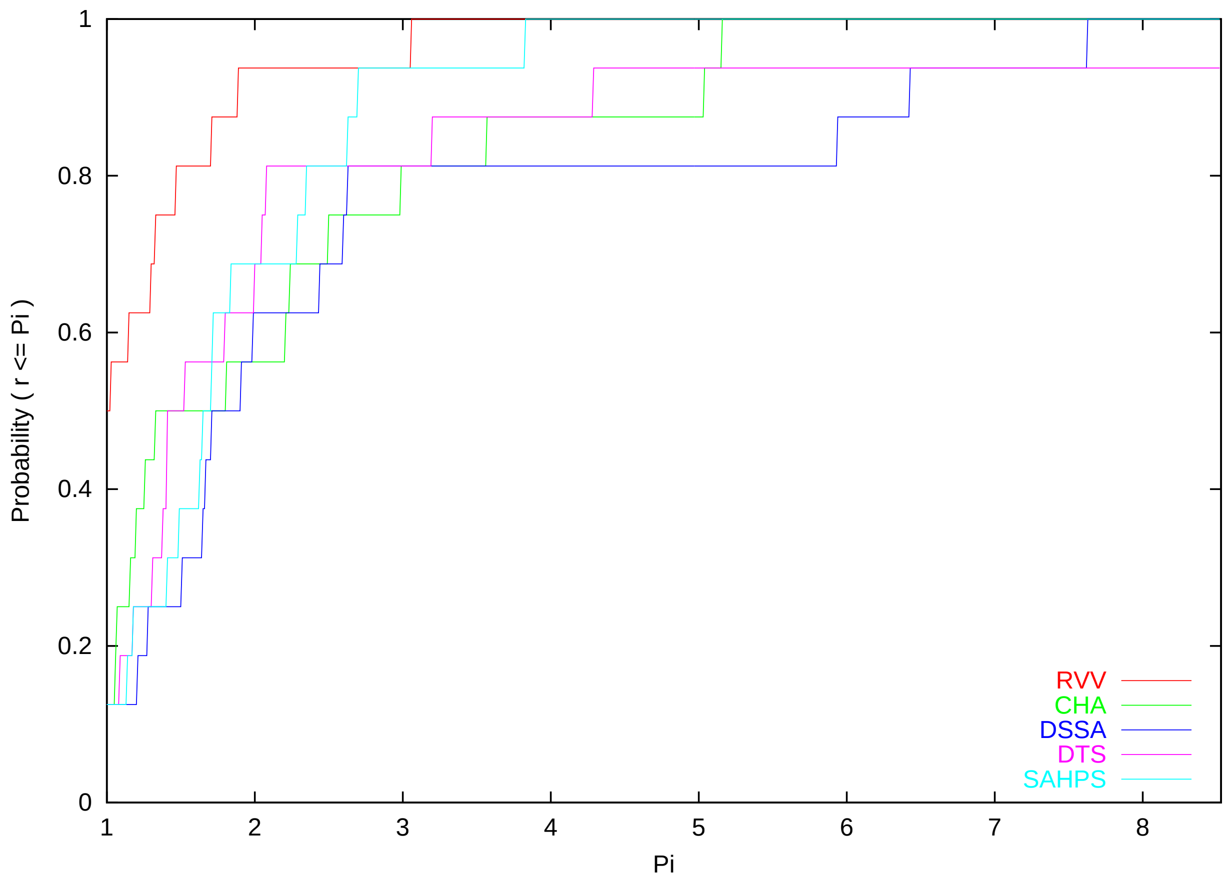
Problem	RVV	CHA (2003)	DSSA (2002)	DTS (2006)	SAHPS (2004)
RC	173	295	118	212	318
ES	189	952	1442	223	432
RT	249	132	252		346
SH	468	345	457	274	450
DJ	104	371	273	446	398
$H_{3,4}$	220	492	572	438	517
$H_{6,4}$	875	930	1737	1787	997
$S_{4,5}$	584	698	993	819	1073
$S_{4,7}$	636	620	932	812	1059
$S_{4,10}$	604	635	992	828	1035
R_2	776	459	306	254	357
Z_2	246	215	186	201	276
R_5	1435	3290	2685	1684	1104
Z_5	820	950	914	1003	716
R_{10}	2827	14563	16785	9037	4603
Z_{10}	1947	4291	12501	4032	2284
HM	290		225		
GR	854		1830		
CV	1344		1592		

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Profile de performance

- 16 problèmes
- Nombre d'évaluations de f

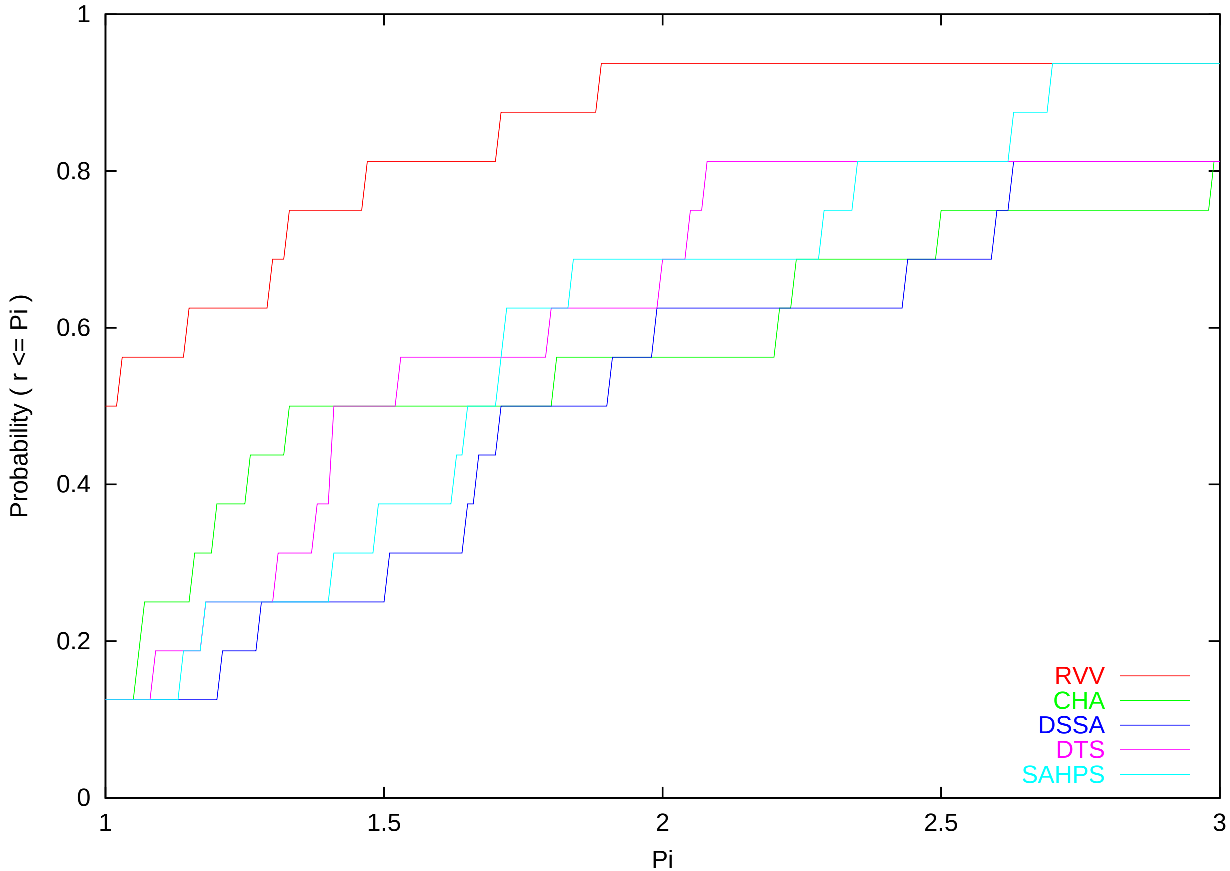


TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Profile de performance: zoom

- 16 problèmes
- Nombre d'évaluations de f



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Conclusions

- Algorithme pour l'optimisation non-linéaire globale en vue d'estimer des modèles de choix discret
  - Nouvelle RVV pour l'optimisation continue
  - Utilisation intensive de l'information sur f et ses dérivées approximées
  - Recherche locale et voisinages "intelligents"
  - Collaboration entre optimisation non-linéaire et optimisation discrète
- Résultats numériques très intéressants
  - La RVV permet de bien diversifier et explorer: bonne robustesse
  - Le coût de la recherche locale compensé par son efficacité

Perspectives

- Tests supplémentaires
  - Poursuivre le paramétrage de l'algorithme
  - Tests en grande taille
  - Tests sur l'estimation de modèles de choix discret
  - Évaluer la performance obtenue pour un budget de temps ou de coût donné
- Développements algorithmiques
  - Utiliser une méthode de filtre-région de confiance
  - Autres variantes
  - Généralisation à l'optimisation sous contraintes

Merci pour votre attention !

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE