Un autre point de vue sur le Lagrangien augmenté pour la gestion de production électrique au court-terme

- Le problème de gestion de production éléctrique à l'horizon journalier vise à fournir
  - pour toutes les unités de production $i \in \{1, 2, \cdots, n \approx 250\}$;
  - sur 96 pas de temps demi-horaires $t \in \{1, 2, \cdots, T = 96\}$;

un programme de production $p \in \mathbb{R}^n \times \mathbb{R}^T$ minimisant la somme des

  - coûts de production séparables en espace (par unité de production) i.e. $C : p \in \mathbb{R}^n \times \mathbb{R}^T \to C(p) = \sum_i C_i(p_i)$;
  - pénalités d'écart à la demande $d \in \mathbb{R}^T$ séparables en temps (par pas de temps) i.e. $\hat{C} : p_d \in \mathbb{R}^T \to \hat{C}(p_d) = \sum_t \hat{C}_t(p_{d,t})$ avec $p_{d,t} = d_t - \sum_i p_{i,t}$;

en satisfaisant les contraintes de fonctionnement $p_i \in X_i$ de chaque unité i.

- Mathématiquement le problème (primal) se formule de la façon suivante :

$$(P) \begin{cases} \min_p \varphi(p) \,, \qquad \text{où} \quad \varphi(p) = \underbrace{\hat{C}(d - \sum_i p_i)}_{\text{Pénalité de défaillance}} + \underbrace{C(p)}_{\text{Coût de fonctionnement}} \\ p = (p_{i,t}) \in X_i \quad \text{programme satisfaisant les contraintes de fonctionnement } X_i \end{cases}$$

- On note P^* l'ensemble des solutions du problème (P).

La phase 1 d'Apogée repose sur l'introduction d'un groupe fictif de défaillance et la dualisation de la contrainte d'offre-demande associée.

• On introduit la variable de défaillance $p_d = d - \sum_i p_i$ et on considère le problème avec contrainte d'égalité offre-demande suivant :

$$\begin{cases} \min_{p, p_d} \tilde{\varphi}(p, p_d) \qquad \text{avec} \quad \tilde{\varphi}(p, p_d) = \hat{C}(p_d) + \sum_i C_i(p_i) \\ p_{d,t} = d_t - \sum_i p_{i,t} \quad \text{contrainte d'égalité offre-demande} \\ p = (p_{i,t}) \in X_i \quad \text{programme satisfaisant les contraintes de fonctionnement } X_i \end{cases}$$

• On dualise la contrainte d'égalité offre-demande en introduisant le Lagrangien simple $L : \mathbb{R}^{n \times T} \times \mathbb{R}^T \times \mathbb{R}^T \mapsto \mathbb{R}$:

$$L(p, p_d, \lambda) = \hat{C}(p_d) + \sum_i C_i(p_i) + \langle \lambda, d - p_d - \sum_i p_i \rangle \ .$$

• La fonction duale $W : \mathbb{R}^T \mapsto \mathbb{R}$ se calcule facilement car naturellement décomposée par unités de production

$$W(\lambda) = \min_{p, p_d} L(p, p_d, \lambda) \ .$$

• Le problème dual (D) consiste à rechercher les multiplicateurs $\lambda \in \mathbb{R}^T$ maximisant la fonction duale (concave par construction) :

$$\max_{\lambda} W(\lambda) \qquad (D)$$
$$D^* := \mathrm{argmax}_\lambda W(\lambda)$$

• Une méthode d'optimisation non-différentiable, la méthode des faisceaux [11] est utilisée pour résoudre le problème dual : on constate une bonne convergence de l'algorithme.

• Cependant,

- Sans convexité du problème primal (P) il existe (a priori) un saut de dualité i.e.

$$\hat{C}(d - \sum_i p_i^*) + \sum_i C_i(p_i^*) - W(\lambda^*) \geq 0 \;, \quad \text{pour tout } (p^*, \lambda^*) \in P^* \times D^* \;.$$

- Sans en plus la différentiabilité de W en $\lambda^* \in D^*$, on ne peut garantir $P^* = \arg\min_{p,p_d} L(p, p_d, \lambda^*)$.

La phase 2 d'Apogée repose sur le dédoublement des programmes et la dualisation par Lagrangien augmenté des contraintes d'égalité des variables dupliquées

- Dédoublement des variables en :
  - programmes statiques n'intervenant que dans la fonction de pénalité $\hat{C}$;
  - programmes dynamiques n'intervenant que dans les coûts de fonctionnement C_i et respectant les contraintes de fonctionnement X_i ;

$$\begin{cases} \min\limits_{p,\hat{p}} \hat{C}(d - \sum_i \hat{p}_i) + C(p) \\ s.c.\ p = \hat{p} \\ p = (p_{i,t}) \in X_i \quad \text{satisfait les contraintes dynamiques} \ . \end{cases}$$

- Principe de lagrangien augmenté dualisant la contrainte d'égalité entre programmes dynamiques et statiques, $p = \hat{p}$:

$$L(p, \hat{p}, \mu, c) = \hat{C}(d - \sum_i \hat{p}_i) + C(p) + \langle \mu, p - \hat{p} \rangle + \underbrace{\frac{c}{2} \|p - \hat{p}\|^2}_{\text{Terme d'augmentation}} \quad .$$

- Dans le problème dual, on se contente de maximiser la fonction duale en μ :

$$\max_{\mu} W(\mu, c) \ , \quad \text{avec } W \text{ la fonction duale} \quad W(\mu, c) = \min_{p,\hat{p}} L(p, \hat{p}, \mu, c) \ .$$

- L'agorithme d'Uzawa est utilisé pour la maximisation en μ de W (différentiable).
- La fonction duale W se calcule difficilement car les termes statiques et dynamiques sont couplés par le terme quadratique d'augmentation du Lagrangien.
  - Le Principe du Problème Auxiliaire (PPA) est utilisé pour découpler p et $\hat{p}$.
  - Les itérations d'Uzawa sont alternées avec celles du PPA.

$$
\left\{
\begin{array}{l}
\mu_{i,t}^{0} = \lambda_t \qquad \text{Initialisation aux multiplicateurs de la 1ère phase} \\[2ex]
\min\limits_{p\,,\hat{p}} \left(
\begin{array}{l}
\sum\limits_{t} \hat{C}_t\left(d_t - \sum_{i \in I} \hat{p}_{i,t}\right) + \cdots \\[2ex]
\sum\limits_{i\,,\,t} \Big[C_{i,t}(p_{i,t}) + \mu_{i,t}(p_{i,t} - \hat{p}_{i,t}) + \overbrace{c(p_{i,t} - \hat{p}_{i,t})(p_{i,t}^{k} - \hat{p}_{i,t}^{k})}^{\text{Linéarisation du Lagrangien augmenté}} + \cdots \\
\qquad \cdots \qquad \underbrace{+ \frac{K}{2}(p_{i,t} - p_{i,t}^{k})^2 + \frac{K}{2}(\hat{p}_{i,t} - \hat{p}_{i,t}^{k})^2}_{\text{Termes de freinage}} \Big]
\end{array}
\right) \Rightarrow (p_{i,t}^{k+1}, \hat{p}_{i,t}^{k+1}) \\[2ex]
\mu_{i,t}^{k+1} = \mu_{i,t}^{k} + \rho(p_{i,t}^{k+1} - \hat{p}_{i,t}^{k+1}) \qquad \text{Multiplicateurs individualisés}
\end{array}
\right.
$$

- A chaque itération k de l'algorithme (Uzawa-PPA), on résout les problèmes statiques et dynamiques suivants :

1. A chaque instant, un problème statique

$$\min_{\hat{p}_{;t}} \hat{C}_t(d_t - \sum_{i \in I} \hat{p}_{i,t}) + \sum_i \big(-\mu_{i,t} - c(p^k_{i,t} - \hat{p}^k_{i,t}) - K\hat{p}^k_{i,t} + \frac{K}{2}\hat{p}_{i,t}\big)\hat{p}_{i,t}$$

2. Pour chaque unité de production i, un problème dynamique

$$\min_{p_{i,.} \in X_i} \sum_t \Big[C_{i,t}(p_{i,t}) + \big(\mu_{i,t} + c(p^k_{i,t} - \hat{p}^k_{i,t}) - Kp^k_{i,t} + \frac{K}{2}p_{i,t}\big)p_{i,t} \Big]$$

3. Mise à jour des multiplicateurs par Uzawa

$$\mu^{k+1}_{i,t} = \mu^k_{i,t} + \rho(p^{k+1}_{i,t} - \hat{p}^{k+1}_{i,t})$$

- Dans le cas convexe, l'utilisation du Lagrangien augmenté rend différentiable la fonction duale et permet ainsi :
  - La stabilité du Lagrangien ou encore la propriété de pénalisation exacte pour tout les points $\mu^* \in D^*$ i.e. en tout point μ^*, maximisant en μ W,

$$\arg\min_{p,\hat{p}} L(p, \hat{p}, \mu^*, c) = P^* \ ,$$

    le calcul de la fonction duale fournit exactement les solutions du problème primal ;
  - L'utilisation d'un algorithme d'optimisation différentiable pour la maximisation de W.

- Difficultés en cas de non convexité,
  - avec le Lagrangien simple, possibilité d'un saut de dualité ;
  - Avec le Lagrangien augmenté, sans optimisation du paramètre c, possibilité d'un saut de dualité ;
  - Avec le Lagrangien augmenté, le problème reste couplé par le terme d'augmentation quadratique, or le PPA utilisé pour les séparer suppose la convexité du problème d'optimisation sous-jacent.

Un autre point de vue sur le Lagrangien augmenté pour la gestion de production électrique au court-terme

• Utiliser les résultats récents sur la dualité dans le cas non convexe pour réduire le saut de dualité [17, 9, 16, 15, 14, 6, 12, 18, 13]

- en choisissant une fonction d'augmentation σ adaptée ;
- en optimisant conjointement les paramètres μ et c caractérisant le Lagrangien augmenté.

• Rechercher des algorithmes permettant de maximiser la fonction duale par rapport aux deux variables duales tout en fournissant des suites primales convergentes [8, 5, 7, 13].

• Tirer profit des résultats récents sur les algorithmes de directions alternées [1, 2] (type Gauss Seidel) dans le cadre non convexe, pour décomposer le problème de minimisation pour le calcul de la fonction duale.

Dualité Lagrangienne

• Le problème primal : on considère des fonctions $\varphi : X \to \mathbb{R}$ et $h : X \to H$

$$\min \varphi(x) \quad s.c. \quad h(x) = 0 \ .$$

• On introduit la fonction de paramétrisation, $f : X \times H \to \mathbb{R}$ paramétrant le niveau de contrainte

$$f(x,z) = \varphi(x) + \chi_{h(x)=z} \ .$$

• Fonction perturbation

$$\beta(z) = \inf_{x \in X} f(x,z) \ .$$

• Lagrangien augmenté $L : X \times H' \times \mathbb{R}^+ \to \mathbb{R}$ avec fonction d'augmentation $\sigma : H \to \mathbb{R}^+$ telle que $\sigma(0) = 0$

$$\begin{aligned} L(x,\lambda,c) &= \inf_{z \in H} \{f(x,z) - \langle \lambda, z \rangle + c\sigma(z)\} \\ &= \varphi(x) - \langle \lambda, h(x) \rangle + c\sigma(h(x)) \ . \end{aligned}$$

• Fonction duale $W : H \times \mathbb{R}^+ \to \mathbb{R} \cup \{-\infty\}$

$$W(\lambda,c) = \inf_{x \in X} L(x,\lambda,c) \ .$$

Dualité de Fenchel-Moreau

• On considère la famille de fonctions $\mathcal{F}_{\lambda,c,d}$ définies sur H à valeurs dans $\mathbb{R}$

$$\mathcal{F} = \{f_{\lambda,c,d} \ : \ z \mapsto \langle \lambda, z \rangle - c\sigma(z) + d\}_{(\lambda,c,d)\in H' \times \mathbb{R}^+ \times \mathbb{R}} \ .$$

• On définit $g^{\mathcal{F}}$, l'enveloppe inférieure de g dans la famille $\mathcal{F}$: en chaque point $z \in H$, $g^{\mathcal{F}}(z)$ coincide avec la fonction $f_{\lambda,c,d} \in \mathcal{F}$ minorante de g la plus proche de $g(z)$ en z

$$g^{\mathcal{F}}(z) := \sup_{\lambda,c,d} \left\{ f_{\lambda,c,d}(z) \ : \quad f_{\lambda,c,d}(\tilde{z}) \le g(\tilde{z}) \text{ pour tout } \tilde{z} \in H \right\} \ .$$

• Conjuguée de Fenchel-Moreau $g^c : H' \times \mathbb{R}^+ \to \mathbb{R}$, associée au couplage $\rho : (z, \lambda, c) \to \langle \lambda, z \rangle - c\sigma(z)$

$$g^c(\lambda, c) = \sup_z \{\rho(z, \lambda, c) - g(z)\} \ .$$

Pour des paramètres (λ, c) donnés $-g^c(\lambda, c)$ fournit la constante d maximale telle que $f_{\lambda,c,d}$ soit une minorante de g.

• Biconjuguée de Fenchel-Moreau $g^{cc} \ : \ H \to \mathbb{R}$

$$g^{cc}(z) = \sup_{\lambda,c} \{\rho(z, \lambda, c) - g^c(\lambda, c)\} \ .$$

• Par définition, la bi-conjuguée de Fenchel-Moreau se confond avec l'enveloppe inférieure de g dans la famille de fonctions $\mathcal{F}$: $g^{cc} \equiv g^{\mathcal{F}}$.

Dualité de Fenchel-Moreau et dualité Lagrangienne

• La fonction duale s'identifie à l'opposé de la tranformée de Fenchel-Moreau de la fonction perturbation

$$\begin{aligned} W(\lambda, c) &= \inf_{x \in X} L(x, \lambda, c) \\ &= \inf_{x \in X, z \in H} \{f(x, z) - \langle \lambda, z \rangle + c\sigma(z)\} \\ &= \inf_{z \in H} \{\beta(z) - \langle \lambda, z \rangle + c\sigma(z)\} \\ &= -\beta^c(\lambda, c) \ . \end{aligned}$$

• Le problème dual consiste à calculer en $z = 0$ la biconjuguée de la fonction perturbation

$$\begin{aligned} \max_{\lambda, c} W(\lambda, c) &= \max_{\lambda, c} \{-\beta^c(\lambda, c)\} \\ &= \sup_{\lambda, c} \{\langle \lambda, 0 \rangle - c\sigma(0) - \beta^c(\lambda, c)\} \ , \quad \text{car, par définition} \quad \sigma(0) = 0 \\ &= \beta^{cc}(0) \ . \end{aligned}$$

• Comme la biconjugée est l'enveloppe inférieure de g dans la famille de fonction $\mathcal{F}$, on retrouve l'inégalité de dualité faible

$$\inf_{x \in X, \, h(x) = 0} \varphi(x) = \beta(0) \geq \beta^{cc}(0) = \max_{\lambda, c} W(\lambda, c) \ .$$

La dualité de Fenchel-Moreau fournit une interprétation géométrique utile pour interpréter le saut de dualité. . . et le réduire.

• Résultat de saut de dualité nul [17]
Supposons que

- la fonction d'augmentation $\sigma : X \to \mathbb{R}^+$ est sci convexe telle que

$$\arg\min \sigma(x) = 0 \quad \text{et} \quad \sigma(0) = 0 \ ,$$

- la fonction de paramétrisation f est propre, sci et *level bounded in x locally uniform in z*.
- Supposons qu'il existe $(\lambda, c) \in H \times \mathbb{R}^+$ tel que

$$\inf_{x,z} \{ f(x,z) - \langle \lambda, z \rangle + c\sigma(z) \} > -\infty \ .$$

Alors, le saut de dualité est nul.

Ce résultat a été étendu de mutliples manières dans [17, 9, 16, 15, 14, 6, 12, 18, 13].

[13] propose un algorithme de sous-gradient modifié, MSg améliorant celui de [8] génèrant une suite croissante $W(\lambda_{k+1}, c_{k+1}) > W(\lambda_k, c_k)$ convergeant vers le max W.

Initialisation $k := 0,\ \lambda_0 \in H'$, $c_0 \in \mathbb{R}_+^*,\ \beta \geq \eta > 0,\ \alpha > 0,\ (\alpha_k) \in]0, \alpha[^{\mathbb{N}}$

Iterations

1. Trouver $x_k \in \arg\min_x \{L(x, \lambda_k, c_k)\}$.
2. si $h(x_k) = 0$, STOP.
3. Actualisation des paramètres : $\eta_k = \min(\eta, ||h(x_k)||)$, $\beta_k = \max(\beta, ||h(x_k)||)$
4. Actualisation des multiplicateurs : choisir $s_k \in [\eta_k, \beta_k]$
   - $\lambda_{k+1} = \lambda_k - s_k h(x_k)$
   - $c_{k+1} = c_k + (1 + \alpha_k) s_k ||h(x_k)||$
   - $k := k + 1$

• Si le problème dual possède une solution, alors (λ_k, c_k) est une suite bornée convergeant vers $(\lambda^*, c^*) \in D^*$ et tous les points d'accumulation de la suite primale (x_k) sont solutions du problème primal, sous certaines conditions sur la fonction d'augmentation (vérifiées par exemple pour $\sigma(x) = \|x\|$).

• Une autre version de cet algorithme (MSg2) permet la convergence en un nombre fini d'itérations.

• Une version inexacte de cet algorithme (IMSg) autorise la résolution approchée du calcul de la fonction duale et des x_k associés.

- On veut minimiser la somme de 2 fonctions couplées par une fonction (*régulière*) Q

$$F(x,y) = f(x) + Q(x,y) + g(y) \ , \quad \text{pour tout } (x,y) \in \mathbb{R}^n \times \mathbb{R}^m \ .$$

- Les fonctions f,g et Q sont supposées vérifier les hypothèses de régularité suivantes

$$\begin{cases} F(x,y) = f(x) + Q(x,y) + g(y) \\ f \ : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \ , \ g \ : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\} \text{ fonctions propres lsci} \\ Q \ : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \text{ est une fonction } C^1 \text{ avec } \nabla Q \text{ Lipschitz .} \end{cases}$$

- Algorithme de minimisation alternée (régularisé)

$$\begin{cases} x_{k+1} = \operatorname{argmin} \left\{ F(u, y_k) + \frac{1}{2\lambda_k} \|u - x_k\|^2 \ : \ u \in \mathbb{R}^n \right\} \ , \\ y_{k+1} \in \operatorname{argmin} \left\{ F(x_{k+1}, v) + \frac{1}{2\mu_k} \|v - y_k\|^2 \ : \ v \in \mathbb{R}^m \right\} \ . \end{cases}$$

• Algorithme de minimisation alternée (régularisé)

$$\begin{cases} x_{k+1} = \operatorname{argmin}\left\{F(u, y_k) + \frac{1}{2\lambda_k}\|u - x_k\|^2 \ : \ u \in \mathbb{R}^n\right\} \ , \\ y_{k+1} \in \operatorname{argmin}\left\{F(x_{k+1}, v) + \frac{1}{2\mu_k}\|v - y_k\|^2 \ : \ v \in \mathbb{R}^m\right\} \ . \end{cases}$$

• Attouch etal [1] obtiennent la cvce de l'algorithme, en remplaçant l'hypothèse de convexité par la propriété de Kurdyka-Lojasiewicz de F en tout $\bar{z} \in \partial F(\bar{z})$:
il existe un voisinage U de $\bar{z}$ et $\eta > 0$ tel que pour tout $z \in U \cap [F(\bar{z}) < F < F(\bar{z}) + \eta]$

$$\varphi'(F(z) - F(\bar{z}))\, dist(0, \partial F(z)) \geq c > 0 \ ,$$

avec $\varphi \in C^1([0, \eta), \mathbb{R}^+)$ vérifiant $\varphi(0) = 0$ et $\varphi'(s) > 0$ pour tout $s \in (0, \eta)$.
• Ex : Pour $\varphi(s) = cs^{1-\theta}$ avec $\theta \in [1/2; 1)$ et F différentiable l'inégalité de Kurdyka-Lojasiewicz donne

$$|F(z) - F(\bar{z})|^{\theta} \leq C\|\nabla F(z)\| \ .$$

• Attouch etal [1] obtiennent
- vers les points critiques ;
- vers le minimum global à condition de ne pas partir trop loin du minimum ;
- des vitesses de convergence si on connaît la fonction φ.

References

H. Attouch, J. Bolte, P. Redont, A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems. An approach based on the Kurdyka-Lojasiewicz inequality.

H. Attouch, J. Bolte, B. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods.

D. Bertsekas *Constrained Optimization and Lagrange Multiplier methods.* 1982

F. Bonnans, J-C. Gilbert, C. Lemaréchal, C. Sagastizábal. Optimisation numérique. Aspects Théoriques et Pratiques. Springer-Verlag, 1997.

R. Burachik et al. On a Modified Subgradient Algorithm for Dual Problems via Sharp Augmented Lagrangians. *Journal of Global Optimization.* **24:** 55-78, 2006.

R. Burachik, A. Rubinov. Abstract Convexity and Augmented Lagrangians. *SIAM J. Optim.*, **18**, No. 2, 413-436, 2007.

R. Burachik, A. Iusem, J. Melo. A primal dual modified subgradient algorithm with sharp Lagrangian. *J. Glob. Optim.* **46**: 347-361, 2009.

R. Gasimov. Augmented Lagrangian Duality and Nondifferentiable Optimization Methods in Nonconvex Programming. *Journal of Global Optimization* **24**: 187-203,

- R. Gasimov, A. Rubinov. On augmented Lagrangians for Optimization Problems with a Single Constraint. *Journal of Global Optimization* **28**: 153-173, 2004.
- X. Huang, X. Yang. A Unified Augmented Lagrangian Approach to Duality and Exact Penalization. *Mathematics of Operations Research* **28**, No. 3, 533-552, 2003.
- C. Lemarechal, A.S. Nemirovskii and Yu. E. Nesterov. New variants of bundle methods. *Mathematical Programming*, 69, 111-147, 1995.
- Q. Liu, X. Yang, H. W. J. Lee. On Saddle Points of a Class of Augmented Lagrangian Functions. *Journal of Industrial and Management Optimization*. 693-700, 2007.
- J. Melo *On General Augmented Lagrangians and a Modified Subgradient Algorithm,* Thèse de Doctorat. Instituto Nacional de Matematica Pura e Aplicada, Brésil.
- A. Nedić, A. Ozdaglar, A. Rubinov. Abstract convexity for nonconvex optimization duality. *Optimization* **56**, 655-674, 2007.
- A. Nedić, A. Ozdaglar. A Geometric Framework for Nonconvex Optimization Duality using Augmented Lagrangian Functions. *Journal of Global Optimization* **40**, 2008
- J.-P. Penot, A. Rubinov. Multipliers and general Lagrangians. *Optimization* **54**,443-467 2005.
- Rockafellar, R.T. and Wets, R.J.-B. Variational analysis. Springer, Berlin, (1998).
- C. Wang, J. Zhou, X. Xu. Saddle Points Theory of Two Classes of Augmented