

Lagrangiens et Pénalités augmentés

Jean-Pierre Dussault

Département de mathématiques
et d'informatique
Université de Sherbrooke

Problèmes

$$\text{(PEQ)} \quad \begin{cases} \min_{x \in \mathbb{R}^n} f(x) \\ \text{sujet à } g(x) = 0 \\ g : \mathbb{R}^n \longrightarrow \mathbb{R}^m, m \leq n \end{cases}$$

$$\text{(PINEQ)} \quad \begin{cases} \min_{x \in \mathbb{R}^n} f(x) \\ \text{sujet à } g(x) \leq 0 \\ g : \mathbb{R}^n \longrightarrow \mathbb{R}^m \end{cases}$$

$$\text{(PLIN)} \quad \begin{cases} \min_{x \in \mathbb{R}^n} cx \\ \text{sujet à } Ax \leq 0 \end{cases}$$

Plan

1. **Pénalités et barrières classiques.**
 - (a) (PEQ) : Pénalité quadratique (R. Courant)
 - (b) (PINEQ) : Barrière logarithmique (Frisch)
 - (c) SUMT : Fiacco & McCormick
2. **Critiques et substitutions aux méthodes de pénalité**
 - (a) Mauvais conditionnement
 - (b) (PEQ) : Lagrangiens augmentés
 - (c) Pénalités exactes
3. **(PLIN) : Karmarkar**
4. **Renouveau des pénalités**
 - (a) Comment contourner le mauvais conditionnement
 - (b) Minimisations approchées et extrapolations

(c) (PEQ) : Pénalité augmentée

(d) (PINEQ) : Nouvelles fonctions de pénalités

(e) Autres aspects numériques

(f) exemple

5. **Conclusion**

1 a) (PEQ) : Pénalité quadratique

$$\begin{array}{ll} \min_{x \in \mathbb{R}^n} & f(x) \\ \text{sujet à} & g(x) = 0 \end{array}$$

- Lagrange : condition nécessaire d'optimalité.

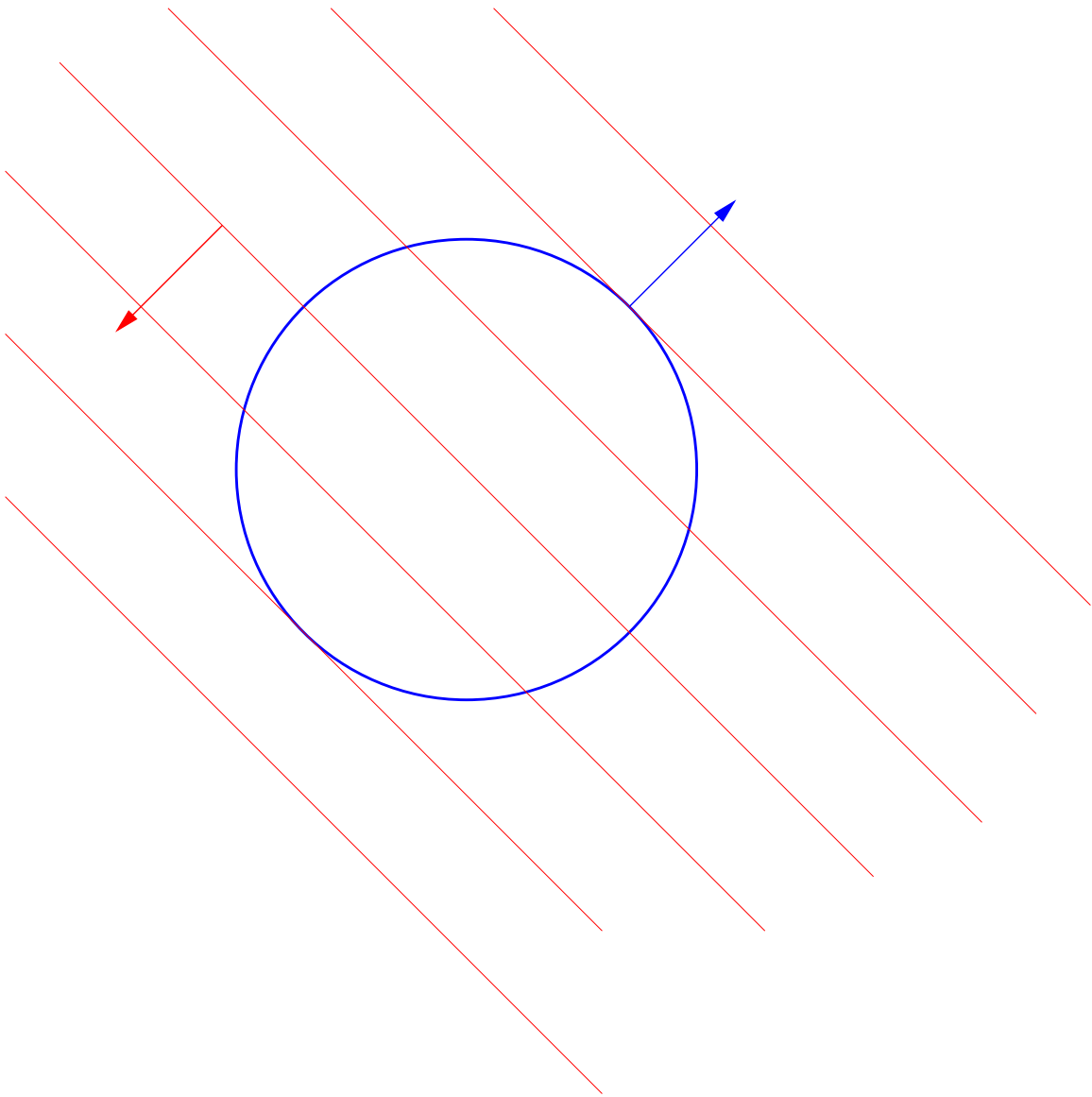
Si x^* est un minimum local pour (PEQ) et x^* régulier pour $g(x^*) = 0$

alors il existe λ^* :

$$\nabla L(x^*, \lambda^*) = \nabla f(x^*) + \lambda^* \nabla g(x^*) = 0$$

$L(x, \lambda) = f(x) + \lambda g(x)$ est nommé Lagrangien.

Preuve Peut être démontré en utilisant la fonction de pénalité !



orthogonalité de ∇f et ∇g

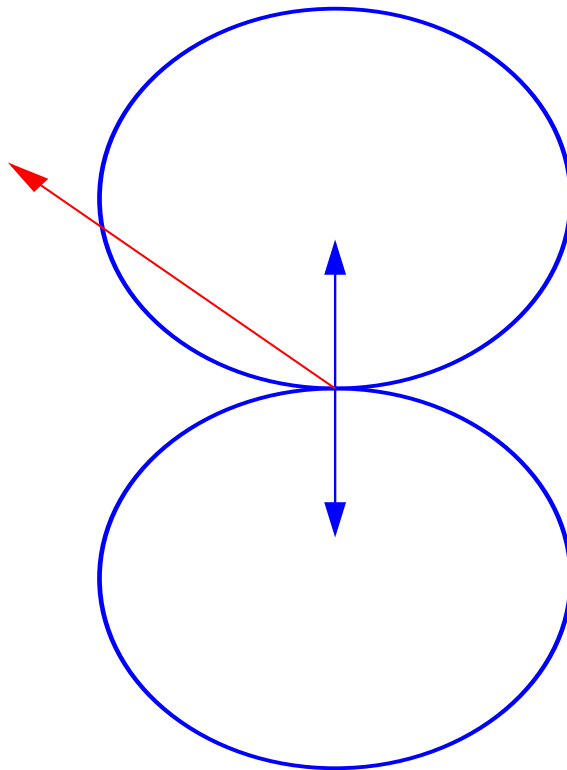
$\min -5(x + y)$ sujet à : $x^2 + y^2 = 1$.

- Qualification des contraintes

x point régulier : $\sum \lambda_i \nabla g_i(x) = 0 \implies \lambda_i \equiv 0$.

Si x^* n'est pas un point régulier de $g(x) = 0$, alors **impasse**.

Exemple : $g(x) = \begin{pmatrix} g_1(x) = x_1^2 + (x_2 - 1)^2 \\ g_2(x) = x_1^2 + (x_2 + 1)^2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.



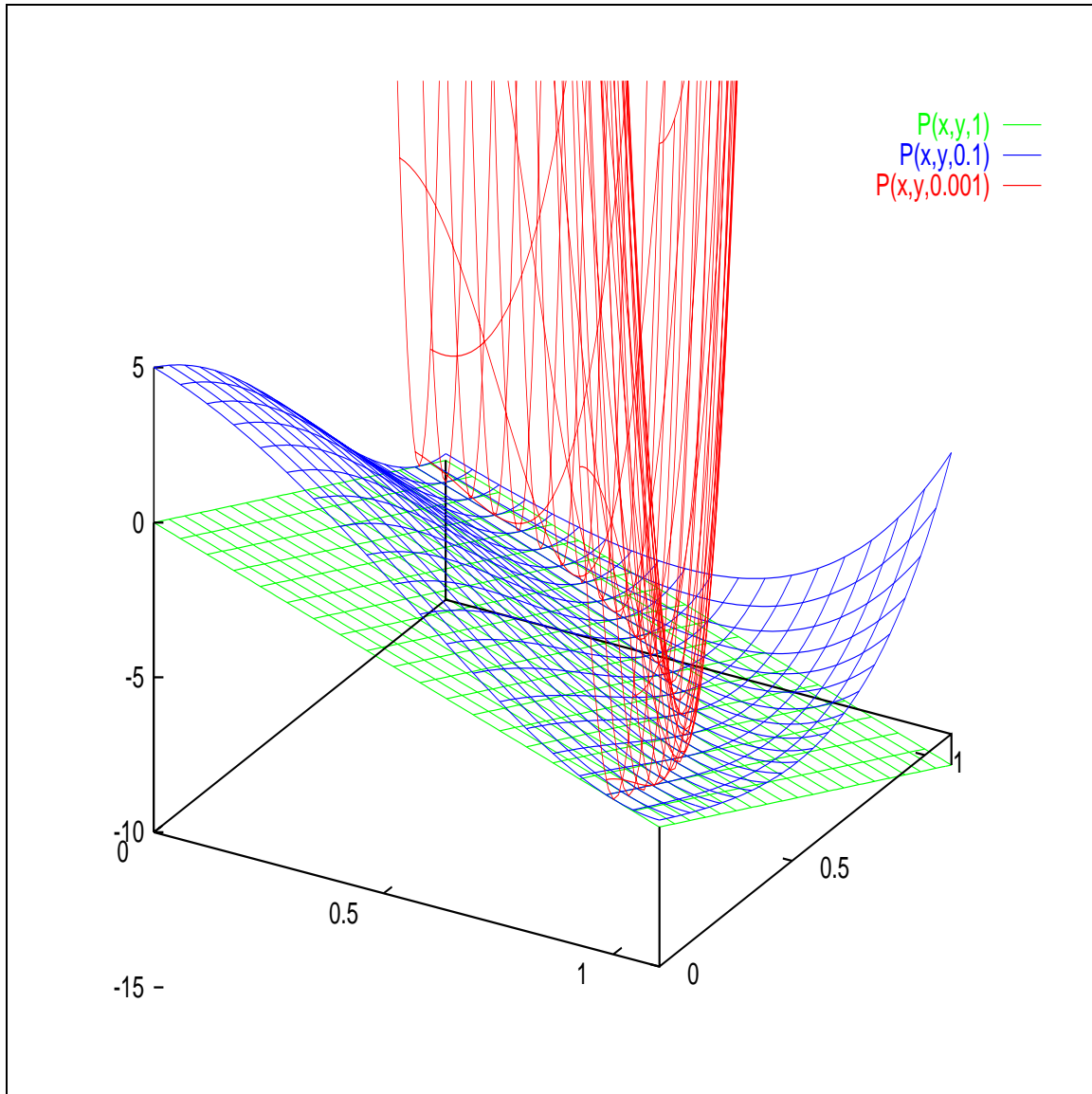
- Pénalité

$$\min_{x \in \mathcal{R}^n} P(x, r) = f(x) + \frac{1}{2r} \|g(x)\|^2$$

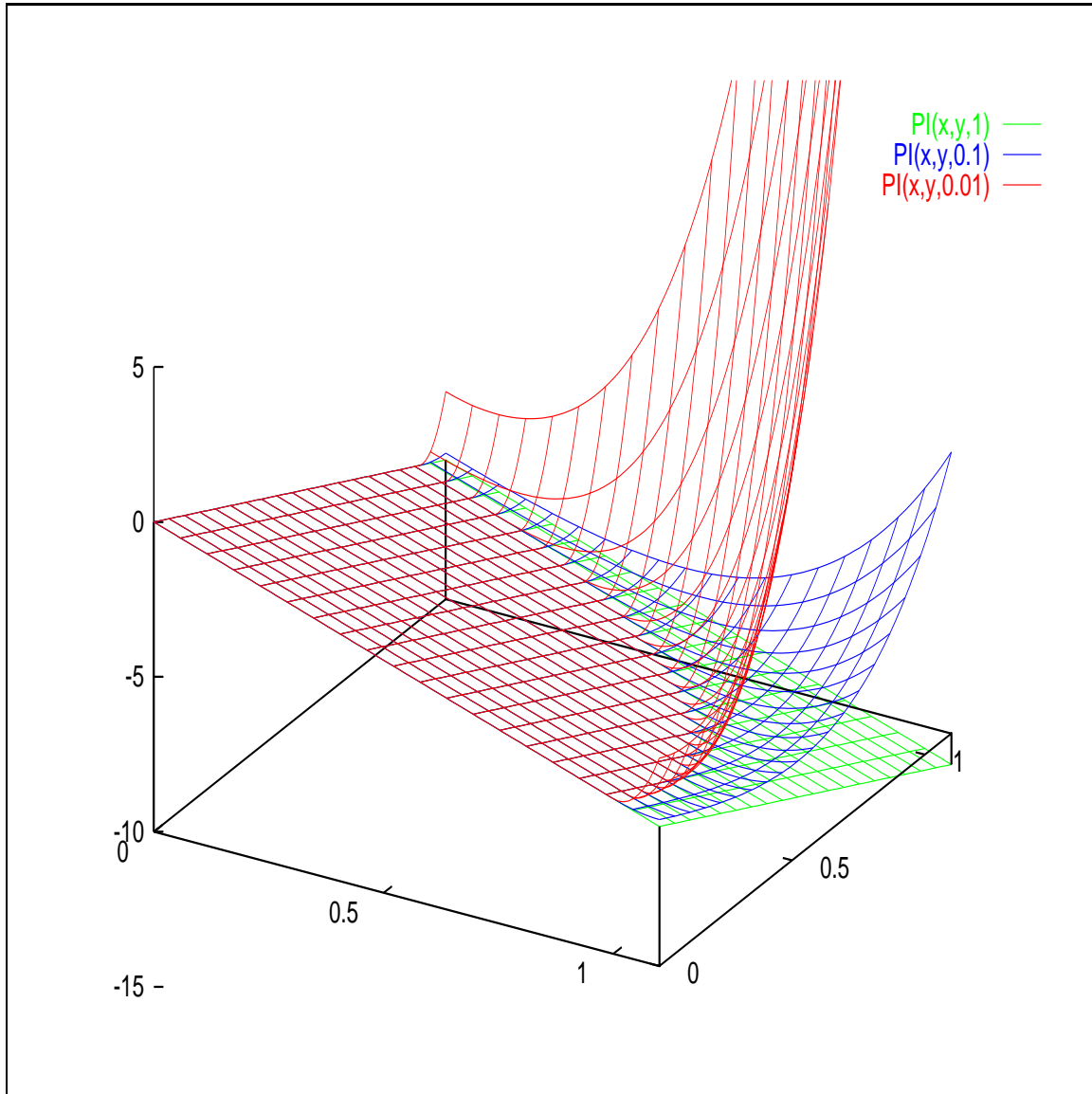
$$\nabla_x P(x, r) = \nabla f(x) + \frac{g(x)}{r} \nabla g(x)$$

Soit une suite $r_k \searrow 0$. Si la suite de points $x_k = x(r_k) : \nabla_x P(x_k, r_k) = 0$ accumule en \bar{x} régulier pour $g(x) = 0$, alors

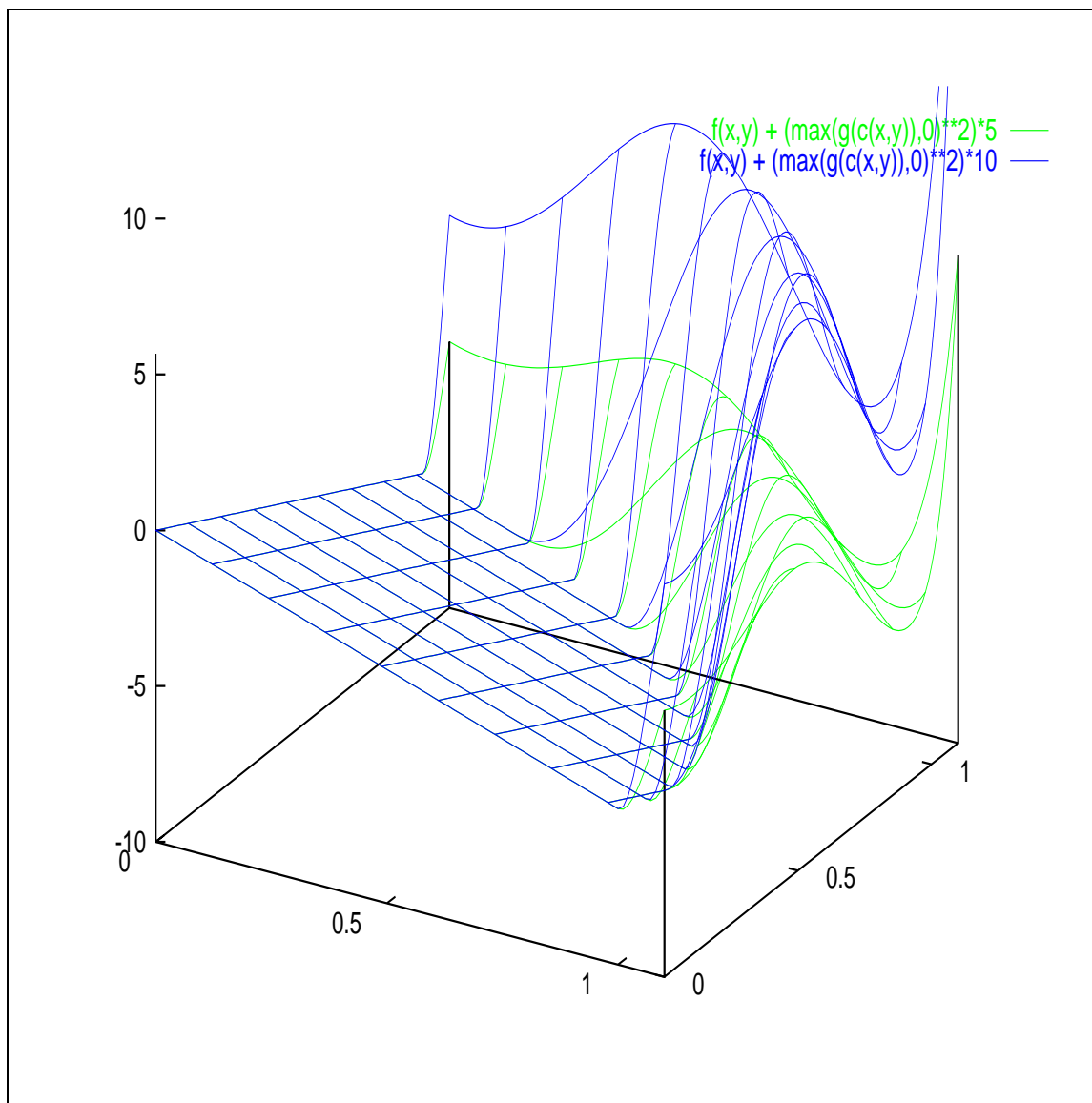
- $\frac{g(x_k)}{r_k}$ accumule en $\bar{\lambda}$
- $\nabla f(\bar{x}) + \bar{\lambda} \nabla g(\bar{x}) = 0$



$$P(x, y, r) = -5(x + y) + 1/2r\|x^2 + y^2 - 1\|^2$$



$$P(x, y, r) = -5(x + y) + 1/2r \|(x^2 + y^2 - 1)_+\|^2$$



$$P(x, y, r) = -5(x + y) + \frac{1}{2r} \|g((x^2 + y^2 - 1)_+)\|^2$$

$$\text{où } g(x) = 4x^3 - 9x^2 + 6x$$

Points minima locaux où $(x^2 + y^2 - 1) > 0$

1 b) (PINEQ) : Barrière logarithmique

$$\begin{aligned} \min_{x \in \mathcal{R}^n} \quad & f(x) \\ \text{sujet à} \quad & g(x) \leq 0 \end{aligned}$$

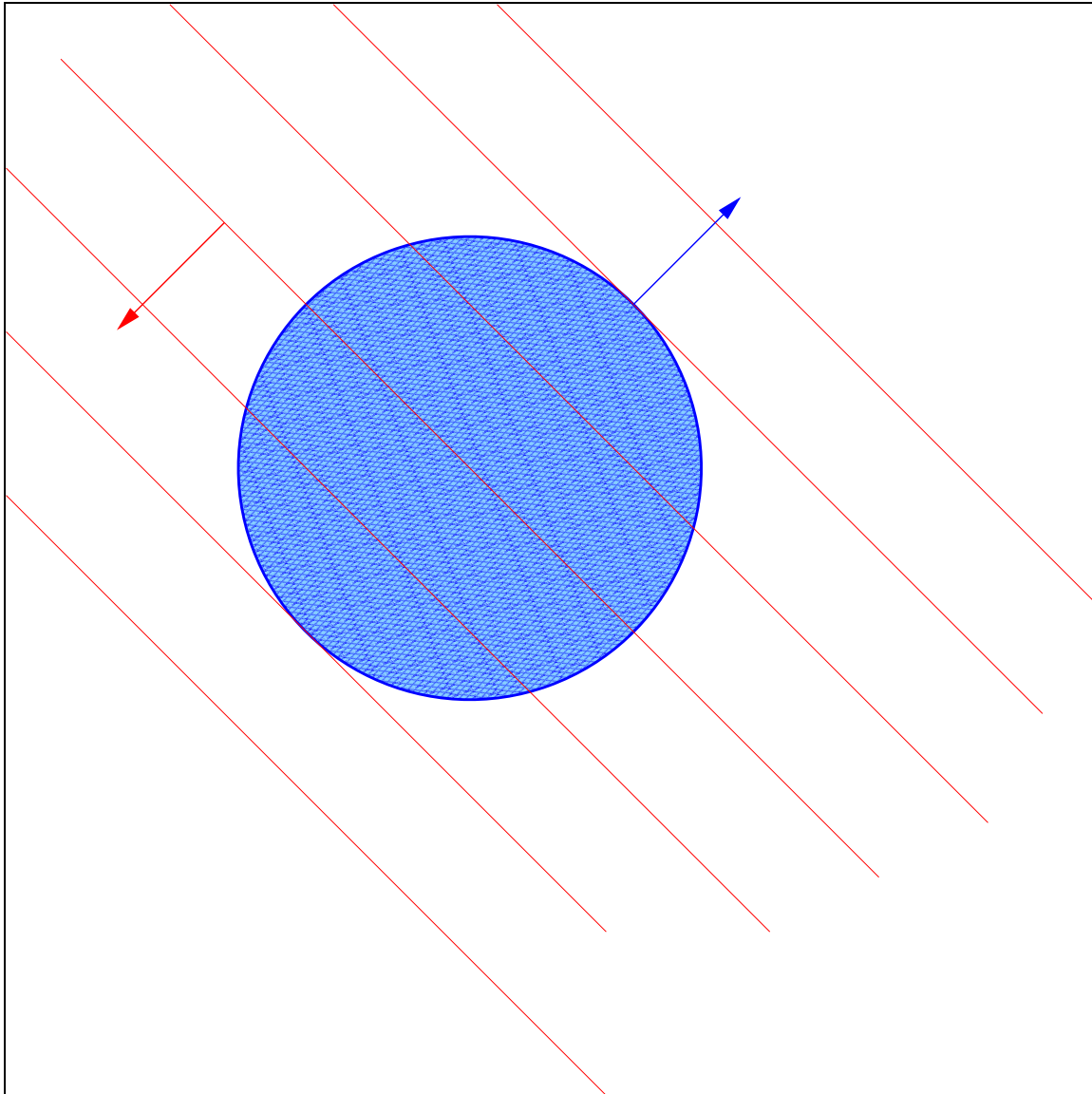
- Karush-Kuhn-Tucker(KKT) :

Si x^* est un minimum local pour (PINEQ) et x^* satisfait une qualification de contraintes pour $g(x^*) \leq 0$

alors il existe $\lambda^* \geq 0$:

$$\begin{aligned} \nabla f(x^*) + \lambda^* \nabla g(x^*) &= 0 \\ \lambda^* g(x^*) &= 0. \end{aligned}$$

Preuve Peut être démontré en utilisant la fonction de pénalité !



orthogonalité de ∇f et ∇g

$\min -5(x + y)$ sujet à : $x^2 + y^2 \leq 1$.

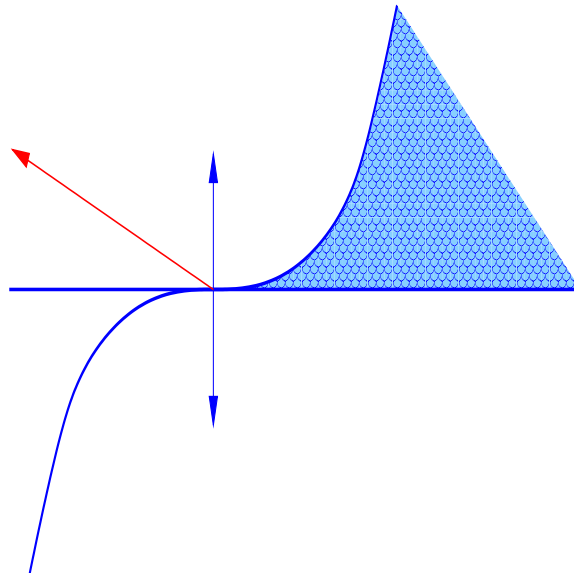
- Qualification des contraintes

x point qualifié :

$$\left\{ \begin{array}{l} \sum \lambda_i \nabla g_i(x) = 0 \\ \lambda \geq 0 \\ \lambda g(x) = 0 \end{array} \right\} \implies \lambda_i \equiv 0.$$

Si x^* n'est pas un point qualifié de $g(x) \leq 0$, alors **impasse**.

Exemple : $g(x) = \begin{pmatrix} g_1(x) = -x_1^3 + x_2 \\ g_2(x) = -x_2 \end{pmatrix} \leq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.



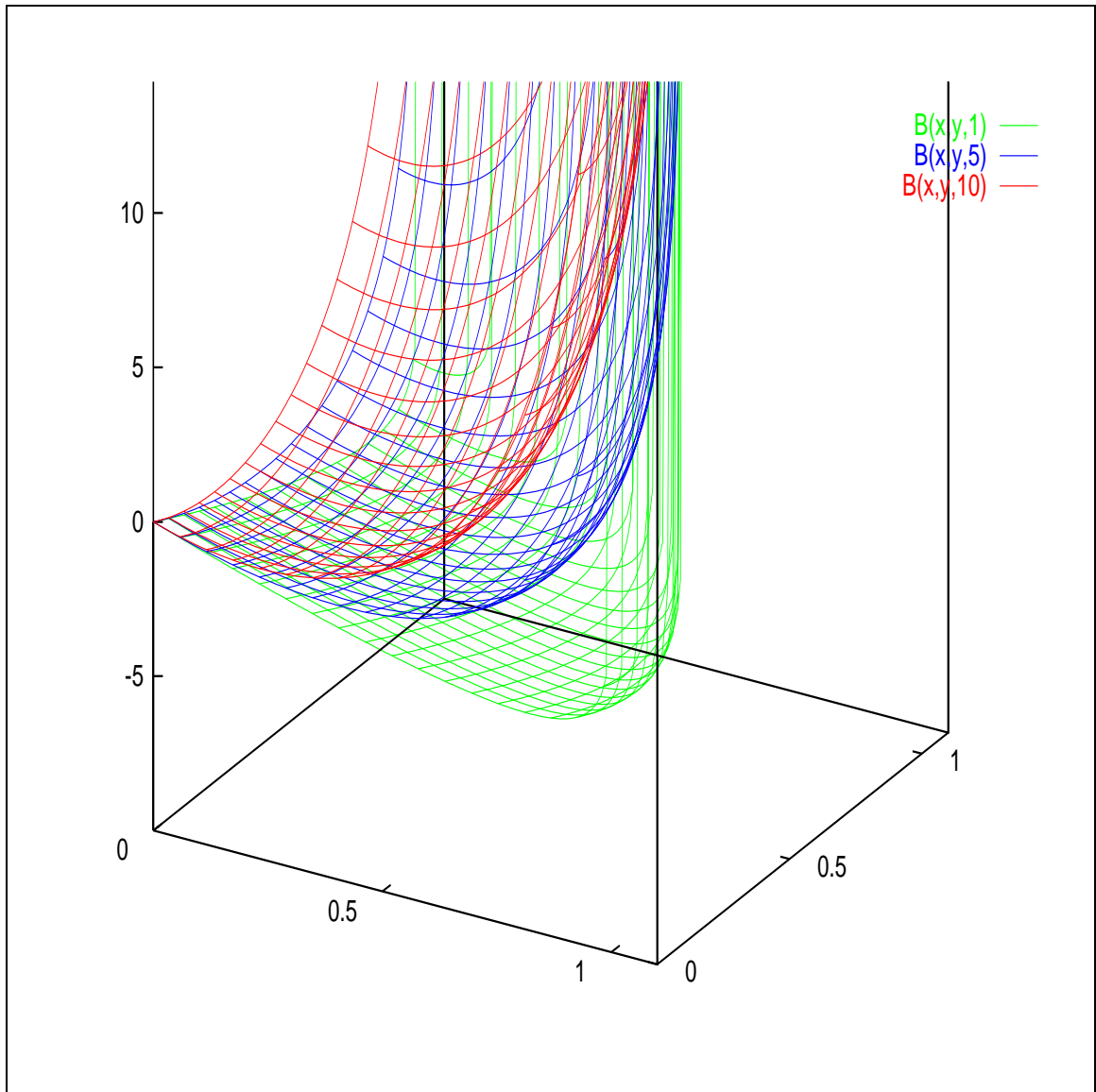
- Barrière

$$\min_{x \in \mathbb{R}^n} B(x, r) = f(x) - r \sum (\log(-g_i(x)))$$

$$\nabla_x B(x, r) = \nabla f(x) - \frac{r}{g} \nabla g(x)$$

Soit une suite $r_k \searrow 0$. Si la suite de points $x_k = x(r_k) : \nabla_x B(x_k, r_k) = 0$ accumule en \bar{x} , un point qualifié,

- $\frac{-r_k}{g(x_k)}$ accumule en $\bar{\lambda} \geq 0$
- $\nabla f(\bar{x}) + \bar{\lambda} \nabla g(\bar{x}) = 0$
- $\bar{\lambda} g(\bar{x}) = 0$.



$$B(x, y, r) = -5(x + y) + r \log(-x^2 - y^2 + 1)$$

1 c) SUMT (PEQ)

$$\nabla_x P(x, r) = \nabla f(x) + \frac{g(x)}{r} \nabla g(x) = 0$$

$$\Theta(x, y, r) = \begin{cases} \nabla f(x) + y \nabla g(x) & = 0 \\ g(x) & = ry \end{cases}$$

Fonctions implicites $x(r)$ et $y(r)$

Condition suffisante d'optimalité

$$\nabla L(x^*, \lambda^*) = 0$$

$$\nabla g(x^*)d = 0 \implies d^t \nabla_{xx}^2 L(x^*, \lambda^*)d > 0$$

x^* point régulier & condition suffisante assure

$\nabla_{xy} \Theta(x, y, 0)$ inversible.

Au voisinage de x^*, λ^* sat. les condition suffisante d'optimalité

$x(r)$ et $y(r)$ constituent des trajectoires isolées, et différentiables en $r = 0$

r petit entraîne que

- $\|x(r) - x^*\| \sim \mathcal{O}(r)$;
- on peut extrapoler sur la trajectoire.

1 c) bis SUMT (PINEQ)

$$\nabla_x B(x, r) = \nabla f(x) + \frac{r}{g(x)} \nabla g(x) = 0$$

$$\Theta(x, y, r) = \begin{cases} \nabla f(x) + y \nabla g(x) & = 0 \\ g_i(x) y_i & = r \end{cases}$$

Fonctions implicites $x(r)$ et $y(r)$

Condition suffisante d'optimalité

$$I^* = \{i : g_i(x^*) = 0\} \quad \bar{I}^* = \{i : \lambda_i^* > 0\}$$

$$\nabla L(x^*, \lambda^*) = 0$$

$$\nabla g_{\bar{I}^*}(x^*)d = 0 \implies d^t \nabla_{xx}^2 L(x^*, \lambda^*)d > 0$$

x^* régulier, condition suffisante et **condition**

de stricte complémentarité $\lambda_{I^*}^* > 0 \iff I^* \equiv \bar{I}^*$

assure

$\nabla_{xy} \Theta(x, y, 0)$ inversible.

Au voisinage de x^*, λ^* sat. les condition suffisante d'optimalité

$x(r)$ et $y(r)$ constituent des trajectoires isolées, et différentiables en $r = 0$

r petit entraîne que

- $\|x(r) - x^*\| \sim \mathcal{O}(r)$;
- on peut extrapoler sur la trajectoire.

2 a) Mauvais conditionnement

$$\nabla_{xx}^2 P(x, r) = \nabla_{xx}^2 L\left(x, \frac{g(x)}{r}\right) + \underbrace{\frac{\nabla g(x) \nabla g(x)^t}{r}}_{\text{diverge quand } r \searrow 0}$$

$$\nabla_{xx}^2 B(x, r) = \nabla_{xx}^2 L\left(x, \frac{r}{g(x)}\right) + \underbrace{\frac{\nabla g(x) Y^2 \nabla g(x)^t}{r}}_{\text{diverge quand } r \searrow 0}$$

où $Y = \text{diag}(r/g_i(x))$.

Murray et Lootsma ont analysé ce phénomène.

2 b) Lagrangiens augmentés (PEQ)

$$L_a(x, \lambda, r) = \underbrace{f(x) + \lambda g(x)}_{L(x, \lambda)} + \frac{1}{2r} \|g(x)\|^2$$

Observations :

- $\forall r > 0$, x^* est un point stationnaire de $L_a(x, \lambda^*, r)$, i.e. $\nabla_x L_a(x^*, \lambda^*, r) = 0$;
- $\exists \bar{r} : \forall r, 0 < r < \bar{r}$, x^* minimum local de $L_a(x, \lambda^*, r)$;
- $\nabla_x L_a(x, \lambda, r) \equiv \nabla_x L(x, \underbrace{\lambda + \frac{g(x)}{r}}_y)$;
- Mise-à-jour de λ : pas besoin que $r \searrow 0$.

2 c) Pénalités exactes

Difficultés liées à $r_k \searrow 0$.

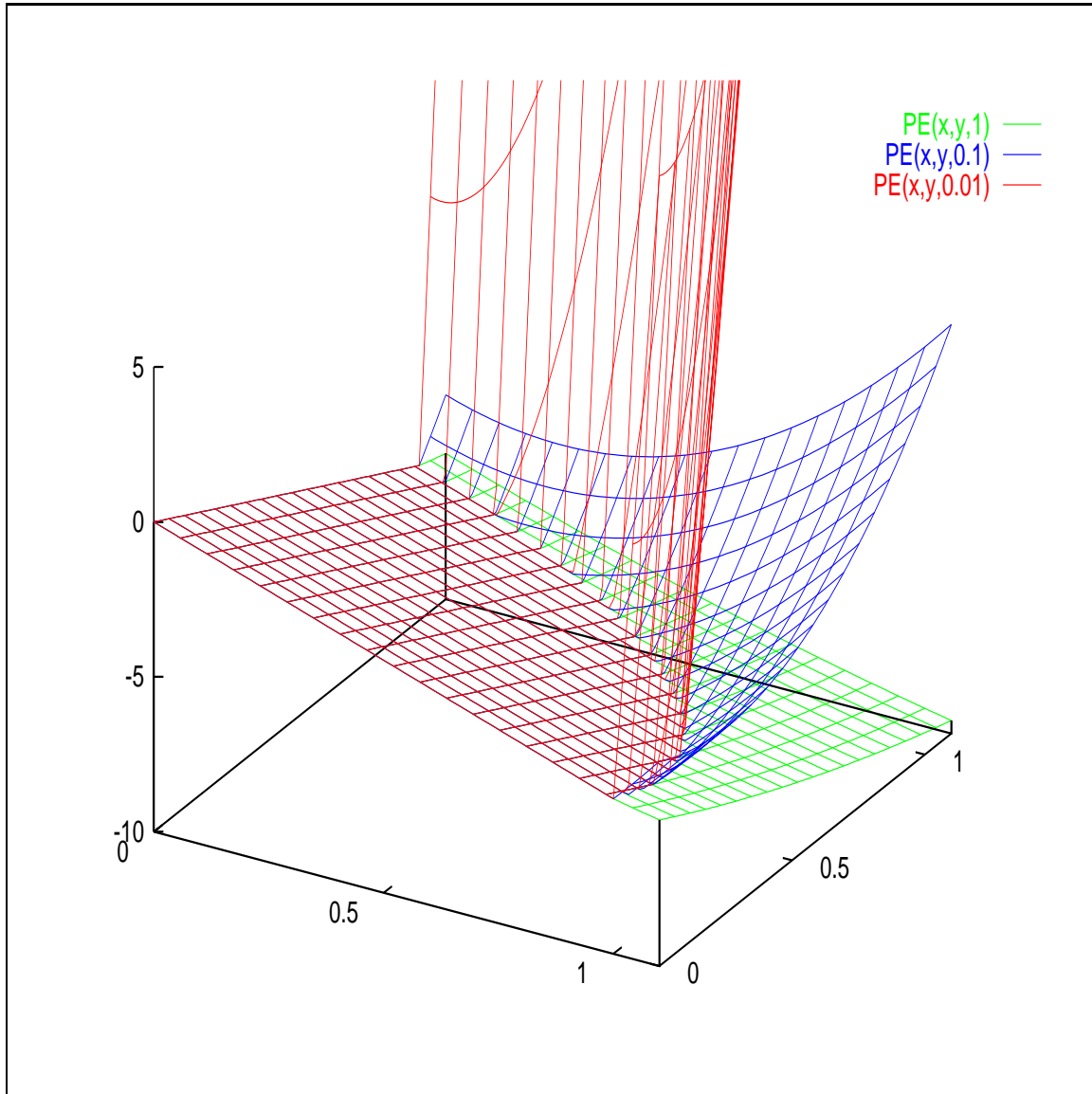
Pénalité qui **ne nécessite pas** $r_k \searrow 0$.

Remplacer $\|g(x)\|^2$ par $\|g(x)\|$.

Minimisation non différentiable.

Il existe des pénalités exactes différentiables complexes.

Utilisée avec succès en combinaison avec les méthodes SQP.



$$PE(x, y, r) = -5(x + y) + 1/r \|(x^2 + y^2 - 1)_+\|$$

3. Karmarkar(PLIN)

Origine des “méthodes de point intérieur”.

Complexité polynômiale.

Implantation efficace meilleure que le simplexe.

Relié à la barrière logarithmique.

4 a) Mauvais conditionnement(PEQ)

Méthode de Newton :

$$d_N = -\nabla_{xx}^2 P(x, r)^{-1} \nabla_x P(x, r)$$

peut être calculée autrement :

$$\Theta(x, y, r) = \begin{cases} \nabla f(x) + y \nabla g(x) \\ g(x) - ry \end{cases}$$

$$\nabla_{xy} \Theta(x, y, r) \begin{pmatrix} d_N \\ \xi \end{pmatrix} = -\Theta(x, y, r)$$

et $\nabla_{xy} \Theta(x, y, r)$ est inversible près de $r = 0$.

4 a) bis Mauvais conditionnement(PINEQ)

Méthode de Newton :

$$d_N = -\nabla_{xx}^2 B(x, r)^{-1} \nabla_x B(x, r)$$

peut être calculée autrement :

$$\Theta(x, y, r) = \begin{cases} \nabla f(x) + y \nabla g(x) \\ g_i(x) y_i - r \end{cases}$$

$$\nabla_{xy} \Theta(x, y, r) \begin{pmatrix} d_N \\ \xi \end{pmatrix} = -\Theta(x, y, r)$$

et $\nabla_{xy} \Theta(x, y, r)$ est inversible près de $r = 0$.

Plusieurs techniques de points intérieurs

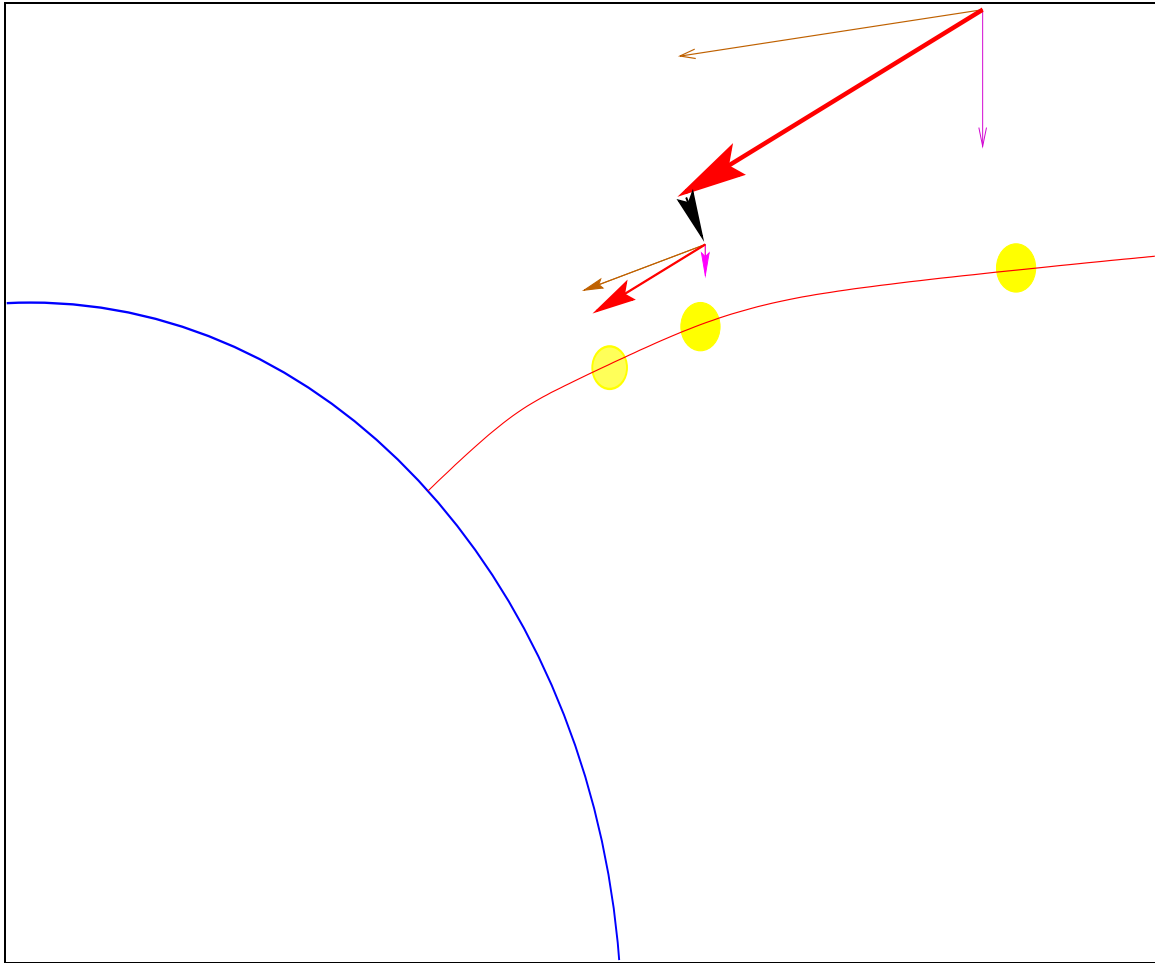
4 b) ϵ -Trajectoires—extrapolations(PEQ)

$$\nabla_x P(x, r) = \psi, \quad \|\psi\| \leq r$$

Au voisinage de x^*, λ^* sat. les condition suffisante d'optimalité

$x(r, \psi)$ constitue une trajectoire isolée, et différentiables en $r = 0, \psi = 0$

$$\hat{x} = x(r, \psi) + \underbrace{\frac{\partial x}{\partial r}(r^+ - r)}_{\text{Extrapolation en } r} + \underbrace{\frac{\partial x}{\partial \psi}(0 - \psi)}_{\text{Newton}}$$



extrapolation

Propriétés de \hat{x} :

- $\|\hat{x} - x(r^+, 0)\| \sim \mathcal{O}(r^2)$
- $\|\nabla_x P(\hat{x}, r^+)\| \sim \mathcal{O}\left(\frac{r^2}{r^+}\right)$
- $\|\nabla_x P(\hat{x} + d_N, r^+)\| \sim \mathcal{O}\left(\frac{r^4}{r^+}\right)$

Si $r^+ = r^\alpha$ avec $\boxed{\alpha < 2}$, asymptotiquement,

$$\boxed{x(r^+, \psi^+) = \hat{x} + d_N, \text{ avec } \|\psi^+\| \leq r^+.$$

4 b) bis ϵ -Trajectoires—extrapolations(PINEQ)

$$\nabla_x B(x, r) = \psi, \quad \|\psi\| \leq r$$

Au voisinage de x^*, λ^* sat. les condition suffisante d'optimalité

$x(r, \psi)$ constitue une trajectoire isolée, et différentiables en $r = 0, \psi = 0$

$$\|\nabla_x P(\hat{x} + d_N, r^+)\| \sim \mathcal{O}\left(\frac{r^4}{r^{+2}}\right)$$

Si $r^+ = r^\alpha$ avec $\alpha < 4/3$, asymptotiquement,

$$x(r^+, \psi^+) = \hat{x} + d_N, \text{ avec } \|\psi^+\| \leq r^+.$$

4 c) Pénalités augmentées(PEQ)

$$\nabla_x L_a(x, \lambda, r) = \psi, \quad \|\psi\| \leq r^\beta$$

Réduire r comme dans les pénalités, et mettre-à-jour λ comme dans les lagrangiens augmentés.

Au voisinage de x^*, λ^* sat. les condition suffisante d'optimalité

$x(r, \lambda, \psi)$ et $y(r, \lambda, \psi) = \lambda + \left(\frac{g(x)}{r}\right)$ constituent des trajectoires isolées, et différentiables en $r = 0, \lambda = \lambda^*, \psi = 0$

Propriétés de $x(r, \lambda, \psi)$ et $y(r, \lambda, \psi)$:

- $\|x(r, \lambda, \psi) - x^*\| \sim \mathcal{O}(\max(r\|\lambda - \lambda^*\|, \|\psi\|))$
- $\|y(r, \lambda, \psi) - \lambda^*\| \sim \mathcal{O}(\max(r\|\lambda - \lambda^*\|, \|\psi\|))$
- $\|\nabla_x L_a(x(r, \lambda, \psi), y(r, \lambda, \psi), r^+)\| \sim \mathcal{O}\left(\frac{r^\beta}{r^+}\right)$
- $\|\nabla_x L_a(x(r, \lambda, \psi) + d_N, y(r, \lambda, \psi), r^+)\| \sim \mathcal{O}\left(\frac{r^{2\beta}}{r^+}\right)$

Si $r^+ = r^\alpha$ et $\|\psi\| < r^\beta$ avec $\alpha < \frac{2\beta}{\beta+1}$ et $\beta \leq 3$, asymptotiquement,

$$x(r^+, y(r, \lambda, \psi), \psi^+) = x(r, \lambda, \psi) + d_N, \text{ avec } \|\psi^+\| \leq r^+.$$

4 d) Pénalités augmentées(PINEQ)

Fonctions de pénalité :

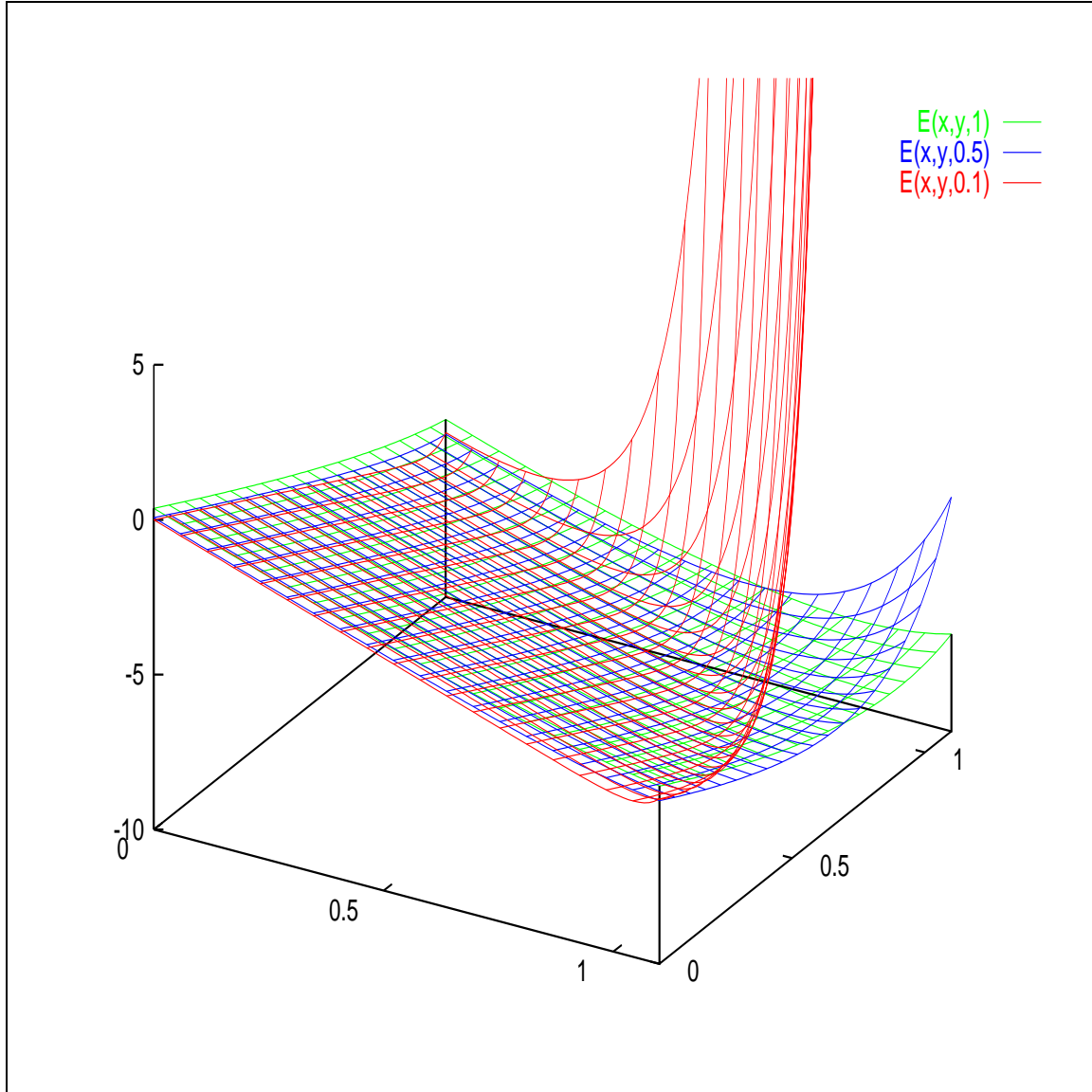
$$P_a(x, \lambda, r) = f(x) + r \sum \lambda_i \psi\left(\frac{g_i(x)}{r}\right)$$

Exemples :

- Pénalité exponentielle : $\psi(t) = e^t$;
- Barrière log modifiée : $\psi(t) = -\ln(1 - t)$;
- Pénalité log—quadratique :

$$\psi(t) = \begin{cases} t + \frac{1}{2}t^2 & \text{if } t \geq -\frac{1}{2} \\ -\frac{1}{4} \log(-2t) - \frac{3}{8} & \text{if } t < -\frac{1}{2} \end{cases}.$$

- Auslender, Cominetti & Haddou (ACH97),
- Ben-Tal & Zibulevsky (BZ97),
- Clovis Gonzaga mercredi.



$$E(x, y, r) = -5(x + y) + r e^{\frac{x^2 + y^2 - 1}{r}}$$

4 e) Autres aspects numériques

Les calculs de $\frac{g(x)}{r}$ ou bien $\frac{r}{g(x)}$.

Exemple : $g(x) = \|x\| - 1$. Précision machine de 16 décimales : $g(x) \sim \mathcal{O}(10^{-8})$ possède seulement 8 décimales de précision relative, et si $r \sim \mathcal{O}(10^{-8})$, $\frac{g(x)}{r}$: 8 décimales absolue.

Constant λ : $g(x) \sim \mathcal{O}(r)$, $\lambda + \frac{g(x)}{r}$ précis jusqu'à $\sqrt{\epsilon_M}$.

Règle de mise-à-jour : $g(x) \sim \mathcal{O}(r^2)$, $\lambda + \frac{g(x)}{r}$ précis jusqu'à $\epsilon_M^{2/3}$.

Pénalité	règle de mise-à-jour	ordre asymptotique	précision limite
Quad	constante	2-steps 2	$\epsilon_M^{1/2}$
Quad	1 ^{er} ordre	1-step 3/2	$\epsilon_M^{2/3}$
ψ, \log	constante	2-steps 4/3	$\epsilon_M^{1/2}$
ψ	1 ^{er} ordre	1-step 4/3	$\epsilon_M^{2/3}$

Puisque λ est mal estimé, changeons-le.

Utilisons à la place un estimateur de moindres carrés :

$$\bar{\lambda} = \arg \min_y \|\nabla f(x_k) + y \nabla g(x_k)\|$$

Utilisons $\bar{\lambda}$ dans $L(x, \lambda)$...

4 f) Un exemple.

$$\begin{array}{ll} \min & \sum_{i=1}^6 ix_i^2 \\ \text{subject to} & (x_1 + x_3 + x_5)^2 = 1 \\ & (x_2 + x_3 + x_4)^2 = 1 \\ & x_1x_6 = 1 \end{array}$$

Resultats pour

- un algo de pénalité ; $\lambda_k \equiv 0$;
- un algo de pénalité augmentée, avec
 1. $r_{k+1} = \min(r_k/5, r_k^{1.1})$, (suite r_k “lente”)
 2. $r_{k+1} = \min(r_k/10, r_k^{1.5})$; (suite r_k “rapide”)
- un algo de Lagrangien augmenté ; $r_k \equiv 0.1$

L'algorithme de pénalité quadratique avec extrapolations.

r_k	iter	$\ \nabla L_a\ $	$\ g\ $	$\nabla_x L$
0.1	1	36.2676	1.29655	
	2	5.51171	0.72136	
	3	2.83569	0.30601	
	4	1.73932	0.57271	
	5	0.23624	0.44846	
	6	0.00405	0.46275	
0.01	extrapol	12.9047	0.10953	
	8	0.47485	0.05317	
	9	0.00044	0.05490	
1D-3	extrapol	1.71108	0.00580	
	11	0.00075	0.00557	
1D-4	extrapol	0.17137	0.00056	
	13	7.4D-07	0.00056	
1D-6	extrapol	0.21590	5.6D-06	
	15	1.4D-08	5.6D-06	
1D-9	extrapol	0.02199	5.6D-09	
	17	1.5D-07	5.6D-09	2D-15
1D-14	extrapol	0.12573	5.6D-14	2.4D-15

L'algorithme de pénalité augmentée avec une suite lente r_k , et une itération de Newton.

r_k	iter	$\ \nabla L_a\ $	$\ g\ $	$\nabla_x L$
0.02	7	42.1698	0.40736	
	8	4.35611	0.04126	
	9	1.89803	0.04551	
	10	0.04803	0.02179	
	11	0.00032	0.02200	
0.004	12	0.13351	0.00025	
	13	0.00028	0.00023	
	14	4.3D-09	0.00023	
0.0008	15	7.9D-05	4.7D-07	
	16	8.3D-10	4.7D-07	
0.00016	17	6.6D-09	2.0D-10	
3.2D-05	18	8.3D-11	1.4D-13	7.5D-11
6.4D-06	19	7.7D-11	2.5D-16	9.5D-16

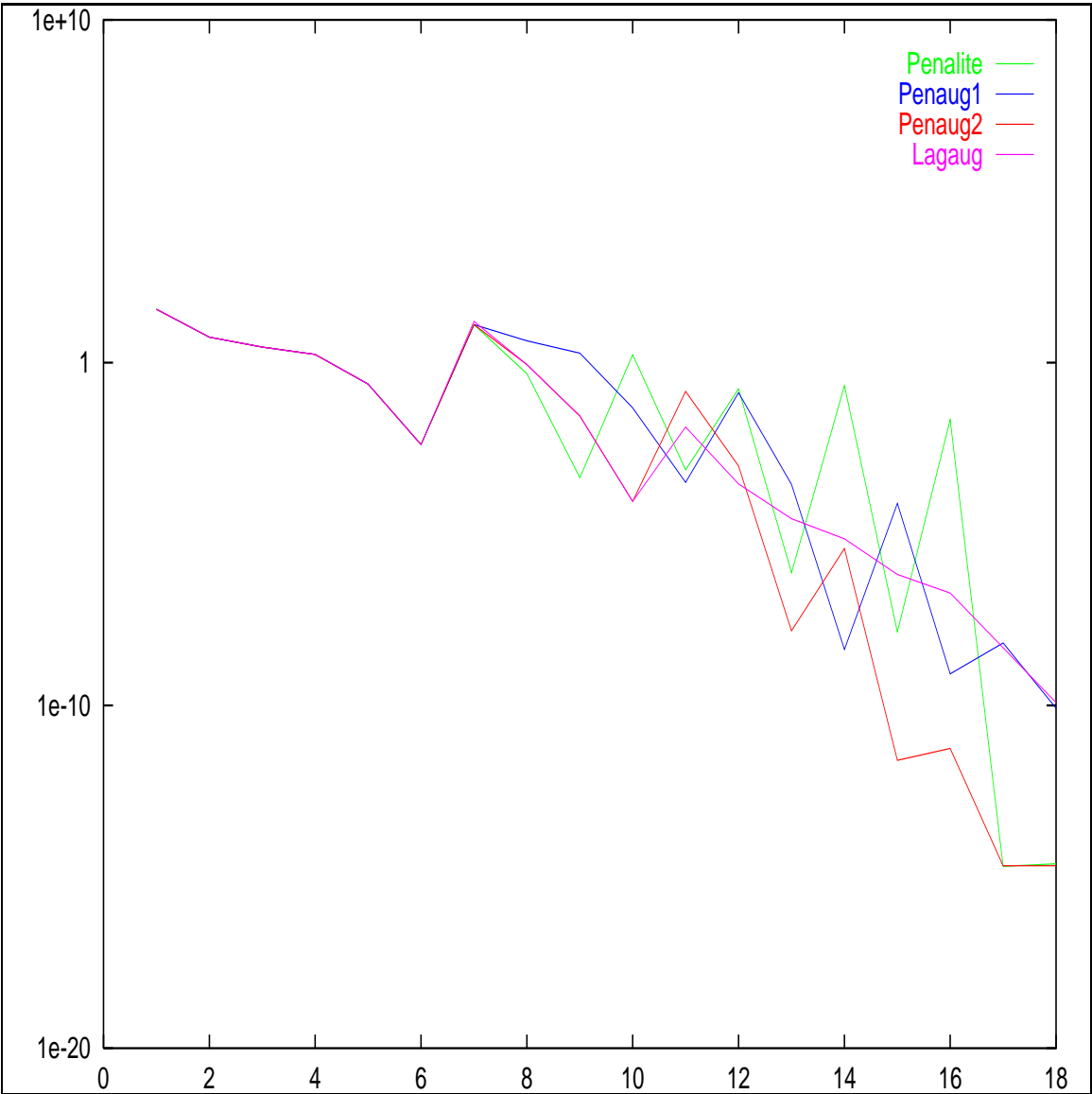
L'algorithme de pénalité augmentée avec une suite rapide r_k , et deux itération de Newton.

r_k	iter	$\ \nabla L_a\ $	$\ g\ $	$\nabla_x L$
0.01	7	15.9853	0.05985	
	8	0.88357	0.01463	
	9	0.02763	0.01141	
	10	8.9D-05	0.01128	
1D-03	11	0.14452	6.1D-05	
	12	0.00097	3.0D-05	
	13	1.5D-08	2.9D-05	
3.2D-05	14	3.4D-05	2.4D-09	3.8D-06
	15	3.3D-09	2.4D-09	2.5D-12
1.8D-07	16	4.5D-09	6.3D-16	5.6D-12
	17	5.4D-09	1.1D-15	2.1D-15

L'algorithme de Lagrangien augmenté

iter	$\ \nabla L_a\ $	$\ g\ $
7	3.91756	0.14905
8	0.40168	0.07565
9	0.00570	0.09141
10	0.05782	0.01805
11	0.00146	0.01932
12	0.00755	0.00406
13	3.6D-07	0.00404
14	0.00140	0.00085
15	4.7D-09	0.00084
16	0.00029	0.00018
17	1.5D-10	0.00017
18	1.0D-05	3.6D-05
19	1.0D-08	7.5D-06
20	4.5D-10	1.6D-06
21	1.9D-11	3.2D-07
22	8.3D-13	6.7D-08
23	3.0D-14	1.4D-08
24	9.3D-15	2.9D-09
...		

... et la convergence linéaire de taux $\sim r = 0.1$.



5 Conclusion

- Propriétés numériques encourageantes.
- Identification efficace des contraintes actives.
- Motivation d'études approfondies.