Support Vector machine
Optimization:Penalty Methods

Professor Abdellatif El Afia

Matematical Model

Consider the following minimization problem:

$$(P)\begin{cases} Min & f(x) \\ s.t & x \in D_R \end{cases}$$

where $D_R = \{x \in \mathbb{R}^n | \boldsymbol{f_i(x) \leq 0} \;\; i \in I_1, \boldsymbol{f_i(x) = 0} \;\; i \in I_2\}$

The Solving of this problem is the solving a suite of unconstraint subproblems, there is three approach as follows:

- **Exterior Penalty Methods:**

  sequence of unconstraint subproblems $(PE)_k$: $Min\{f(x) + r_k PE(x)\} = Min\{E(r_k, x)\} \rightarrow$ **unconstraint optimization**

- **Interior Penalty Methods : Si $I_2 = \{\emptyset\}$**

  sequence of unconstraint subproblems $(PI)_k$: $Min\left\{f(x) + \frac{1}{r_k} PI(x)\right\} = Min\{I(r_k, x)\} \rightarrow$ **unconstraint optimization**

- **Mixed Penalty Methods**

sequence of unconstraint subproblems $(PM)_k$: $Min\{f(x) + M(r_k, t_k)\} = Min\{M(r_k, t_k, x)\} \rightarrow$ **unconstraint optimization**

$$M(r_k, t_k) = r_k PE(x) + \frac{1}{t_k} PI(x)$$

Exterior Penalty Methods

$$(PE)_k\colon\ Min\{f(x) + r_k PE(x)\}$$

The sequence of values $\{r_k\}$ has the following properties:

- Positivity: $\forall k \;\; r_k \geq 0$
- Monotonicity : $\forall k \; r_{k+1} \geq r_k$
- Divergence: $\lim\limits_{k\to\infty} r_k = +\infty$

The Exterior Penalty function is a real value function, $PE\colon \mathbb{R}^n \to \mathbb{R}$, with the following properties:

- PE is a continuous function over $\mathbb{R}^n$
- $\forall x \in \mathbb{R}^n \quad PE(x) \geq 0$
- $PE(x) = 0 \Leftrightarrow x \in D_R(i \in I_1 \; f_i(x) \leq 0, \boldsymbol{i \in I_2} \; f_i(x) = 0)$

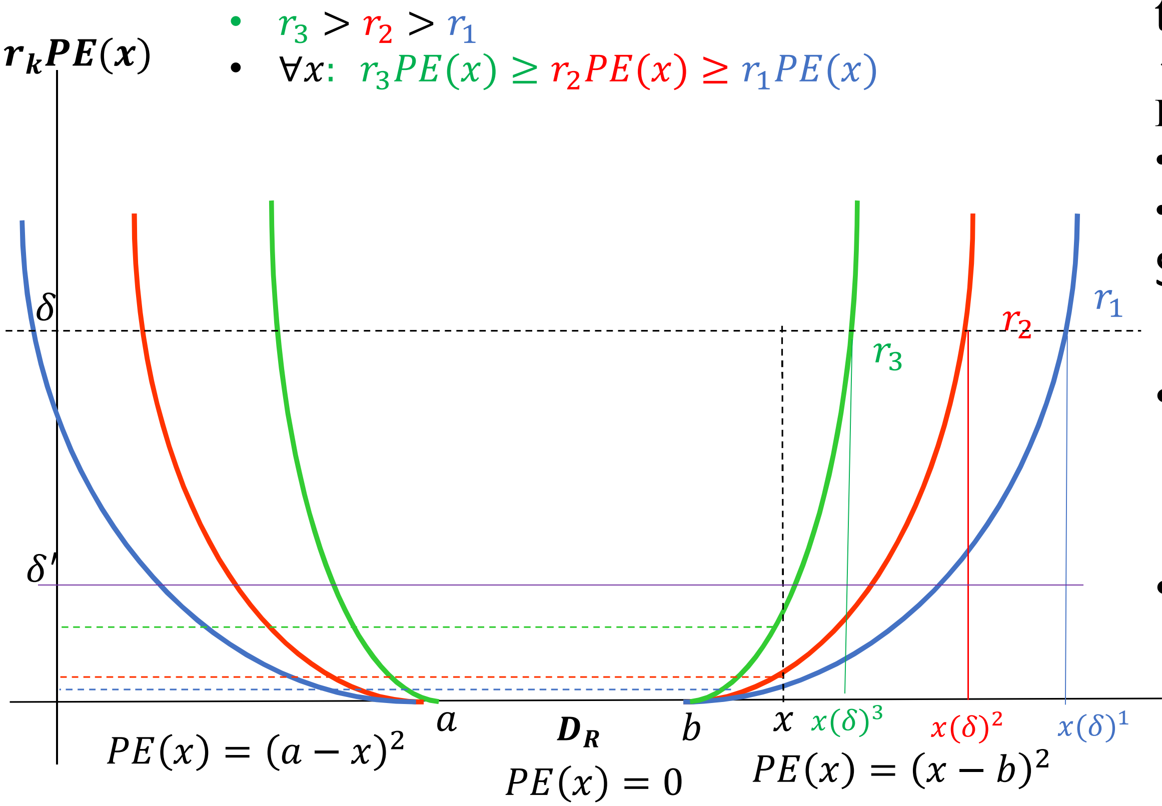
Example:

$$PE(x) = \sum_{i \in I_1} (Max\{0, f_i(x)\})^2 + \sum_{\boldsymbol{i \in I_2}} (\boldsymbol{f_i(x)})^2$$

- $PE(x) = 0 \Leftrightarrow x \in D_R (i \in I_1\ f_i(x) \leq 0, \boldsymbol{i} \in \boldsymbol{I_2}\ f_i(x) = 0)$
- $x \notin D_R (\exists i \in I_1\ f_i(x) > 0, \exists \boldsymbol{i} \in \boldsymbol{I_2}\ f_i(x) \neq 0) \Leftrightarrow PE(x) > 0$

- $PE(x) = \sum_{i \in I_1} (Max\{0, f_i(x)\})^2 + \sum_{\boldsymbol{i \in I_2}} (\boldsymbol{f_i(x)})^2$
- $\boldsymbol{i} \in \boldsymbol{I_2}, f_i(x) = 0 \Leftrightarrow f_i(x) \leq 0, f_i(x) \geq 0$
- $PE(x) = P(f)(x) = \begin{cases} 0 \ \ if \ \ \ f(x) \leq 0 \\ \quad\quad f(x)^2 \end{cases}$
- $PE(x) = Max\{0, f(x)\}$

Exterior Penalty Methods : Example



If $f_1(x) = x - b \leq 0$, $f_2(x) = a - x \leq 0$ then

$rPE(x) = r(Max\{0,(a-x)\})^2 + r(Max\{0,(x-b)\})^2$

Let

- $\delta > 0$,
- the sequence $\{x(\delta)^k\}$ where $r_k PE\left(x(\delta)^k\right) = \delta$

Such that

- if $r_{k+1} > r_k$ then $x(\delta)^{k+1}$ close to $\boldsymbol{D_R}$ than $x(\delta)^k$

- if $\lim_{k\to\infty} r_k = +\infty$ then the solution of $(PE)_k$ converges to the solution of (P)

- $f_1(x) = x - b \leq 0 \Longrightarrow x \leq b$,
- $f_2(x) = a - x \leq 0 \Longrightarrow x \geq a$
- $D_R = [a, b]$
- k: le problème $(PE)_k$
- Soit δ , la suite $\{x(\delta)^k\}$ où $r_k PE\left(x(\delta)^k\right) = \delta$
- $\lim\limits_{k \to +\infty} r_k PE\left(x(\delta)^k\right) = 0 \to \delta \to 0$
- $\lim\limits_{k \to +\infty} (PE)_k = (P) \Leftrightarrow f(x^*) = f(x^*) + r_\infty PE(x^*)\ x^* \in D_R$
- $\lim\limits_{k \to +\infty} r_k PE\left(x^k\right) = 0$
- $PE(x^*) = 0$ car $x^* \in D_R$
- $\sum_{i \in I_1} (Max\{0, f_i(x)\})^2$

Exterior Penalty Methods : Convergence

The next lemme is used to demonstrate the convergence of the method.
Let's denote by x^k the optimal solution of $Min\{E(r_k, x)\}$

Lemme 1 :
for any value of k,

- $E(r_k, x^k) \leq E(r_{k+1}, x^{k+1})$
- $PE(x^k) \geq PE(x^{k+1})$
- $f(x^k) \leq f(x^{k+1})$
- If x^* is an optimal solution of (P) then $\forall k, f(x^*) \geq E(r_k, x^k) \geq f(x^k)$

Theorem 1 :
Let $\{x^k\}$ a sequence of points generated by the Exterior penalty method $(PE)_k$. Any endpoint of this sequence is a solution of (P) .

Exterior Penalty Methods : Algorithm If $I_1 = \{\emptyset\}$

Input :

- $E(r,x) = f(x) + r\sum_{i\in I_2}(f_i(x))^2$
- $x^0, \delta\ , r_0, \varepsilon_0, \alpha > 1, \beta > 1 \qquad \lim_{k\to\infty} \varepsilon_k = 0$
- $\|\nabla E(r_0, x^0)\| \le \varepsilon_0$

- $k = 0$
- WHILE($\|\nabla E(r_k, x^k)\| > \delta$) {
  - $r_{k+1} = (r_k)^\alpha$
  - $\varepsilon_{k+1} = (\varepsilon_k)^\beta\ if\, \varepsilon_0 < 1\ \beta > 1\ else\, \beta < 1$
  - $Min\{E(r_{k+1}, y)\}$
    - $y^0 = x^k$
    - $t = 0$
    - WHILE($\|\nabla E(r_{k+1}, y^t)\| > \varepsilon_{k+1}$){
      - $\boldsymbol{d^k = -\nabla E(r_{k+1}, y^t)}$
      - **Déterminer $\boldsymbol{\alpha_t}$ tel que $\boldsymbol{\alpha_t = \underset{\alpha>0}{argmin}\{f(x - \alpha\nabla E(r_{k+1}, y^t))\}}$ *or inaccurate line search as Armijo rule***
      - $\boldsymbol{y^{t+1} = y^t - \alpha_t \nabla E(r_{k+1}, y^t)}$
      - $\boldsymbol{t = t + 1}$ }
  - $x^{k+1} = y^*\ \|\nabla E(r_{k+1}, x^{k+1})\| \le \varepsilon_{k+1}$
  - $k = k + 1$

  }
- Fin

$x^0 = (0,0)\ r_0 = 2, \delta = 10^{-2}, \alpha = \beta = 2,\ \varepsilon_0 = 22\ \boldsymbol{\left\|\nabla E\left(r_0, x^0\right)\right\| \le \varepsilon_0}$

- $\boldsymbol{MinE(r_0, y) = x^0}$ $, \varepsilon_0 = 0{,}1$ *until* $\boldsymbol{\|\nabla E(r_0, y)\| \le \varepsilon_0}$
- $(P)\begin{cases} Min & \frac{1}{2}(x_1^2 + x_2^2) \\ s.t & x_1 + x_2 = 5 \end{cases}$
- $(PE)_k\colon\ Min\{x_1^2 + x_2^2 + r_k(x_1 + x_2 - 5)^2\}$
- $E(r, x) = x_1^2 + x_2^2 + r(x_1 + x_2 - 5)^2$
- $\rightarrow \nabla E(r_k, x) = \begin{pmatrix} x_1 + 2r_k(x_1 + x_2 - 5) \\ x_2 + 2r_k(x_1 + x_2 - 5) \end{pmatrix}$
- $k = 0$
- $\rightarrow \nabla E(r_0, x^0) = \begin{pmatrix} -20 \\ -20 \end{pmatrix} \rightarrow \|\nabla E(r_0, x^0)\| = \sqrt{400} = 20 > \delta = 10^{-2}$

Interior Penalty Methods

$$(PI)_k:\ Min\left\{f(x) + \frac{1}{r_k} PI(x)\right\} = Min\{I(r_k, x)\}$$

The sequence of values $\{r_k\}$ has the following properties:

- Positivity: $\forall k\ \ r_k \geq 0$
- Monotonicity : $\forall k\ r_{k+1} \geq r_k$
- Divergence: $\lim_{k \to \infty} r_k = +\infty$

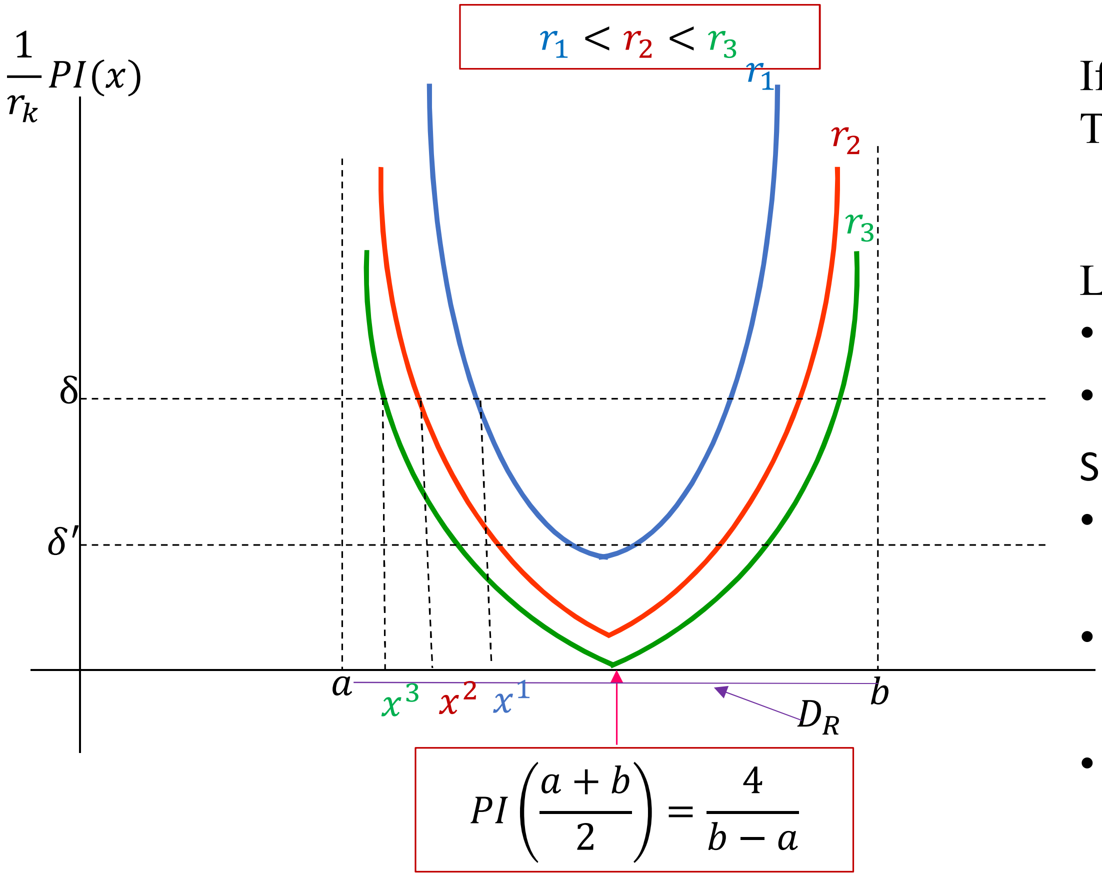
The Interior Penalty function is a real value function, $PI: \mathbb{R}^n \to \mathbb{R}$, with the following properties:

- PI is a continuous function over $\mathbb{R}^n$
- $\forall x \in Int(D_R)\ \ PI(x) \geq 0$
- $\lim_{x \to \bar{x}} PI(x) = +\infty$ when $\bar{x} \in Fr(D_R)$

Example:

- $PI(x) = \sum_{i \in I_1} \frac{-1}{f_i(x)}$
- $PI(x) = -\sum_{i \in I_1} \log\big(-f_i(x)\big)$

Interior Penalty Methods: example



If $f_1(x) = x - b$, $f_2(x) = a - x$
Then

$$PI(x) = -\frac{1}{a-x} - \frac{1}{x-b}$$

Let

- $\delta > 0$,
- the sequence $\{x(\delta)^k\}$ where $\frac{1}{r_k} PI\big(x(\delta)^k\big) = \delta$

Such that:

- if $r_{k+1} > r_k$ then $x(\delta)^{k+1}$ close to $\boldsymbol{D_R}$ than $x(\delta)^k$
- the $\frac{1}{r_k} PI(x)$ curve is closer to the boundaries as x^k increases.
- Note that with each iteration the solution $x^k \in Int(D_R)$ or this is not very restrictive since if we use an iterative method to solve $(PI)_k$, and if the solution $x^{k-1} \in Int(D_R)$ then $x^k \in Int(D_R)$

Interior Penalty Methods: Convergence

The next lemme is used to demonstrate the convergence of the method.
Let's denote by x^k the optimal solution of $Min\{I(r_k, x)\}$

Lemme 2:
for any value of k,

- $I\left(r_k, x^k\right) \geq I(r_{k+1}, x^{k+1})$
- $PI\left(x^k\right) \leq PI(x^{k+1})$
- $f(x^k) \geq f(x^{k+1})$

Theorem 2:
Let $\{\, x^k\}$ a sequence of points generated by the Interior penalty method. Any endpoint of this sequence is a solution of (P) .

Interior Penalty Methods : Algorithm If $I_2 = \{\emptyset\}$

Input :

- $I(r,x) = f(x) + \frac{1}{r}\sum_{i \in I_1} \frac{-1}{f_i(x)}$ Ou $I(r,x) = f(x) - \frac{1}{r}\sum_{i \in I_1} \log\left(-f_i(x)\right)$
- $x^0, \delta\ , r_0, \varepsilon_0, \alpha > 1, \beta > 1\ \lim_{k \to \infty} \varepsilon_k = 0$
- $\|\nabla I(r_0, x^0)\| \leq \varepsilon_0$ solve $x^0 = argmin I(r_0, y)$ unconstraint optimization

- $k = 0$
- WHILE($\left\|\nabla I\left(r_k, x^k\right)\right\| > \delta$) {
  - $r_{k+1} = (r_k)^\alpha,\quad \varepsilon_{k+1} = (\varepsilon_k)^\beta$
  - $Min\{I(r_{k+1}, x)\}$
    - $y^0 = x^k$
    - $t = 0$
    - WHILE($\|\nabla I(r_{k+1}, y^t)\| > \varepsilon_{k+1}$){
      - $\boldsymbol{d^k = -\nabla I(r_{k+1}, y^t)}$
      - **Déterminer $\boldsymbol{\alpha_t}$ tel que $\boldsymbol{\alpha_t = \underset{\alpha>0}{argmin}\{f\left(x - \alpha \nabla I(r_{k+1}, y^t)\right)\}}$ *or inaccurate line search as Armijo's Algorithm***
      - $\boldsymbol{y^{t+1} = y^t - \alpha_t \nabla I(r_{k+1}, y^t)}$
      - $\boldsymbol{t = t + 1}$ }
  - $x^{k+1} = y^*$
  - $k = k + 1$

  }
- Fin

Mixed Penalty Methods

$$(PM)_k:\ Min\left\{f(x) + r_k PE(x) + \frac{1}{t_k} PI(x)\right\} = Min\{M(r_k, t_k, x)\}$$

The sequence of values $\{r_k, t_k\}$ has the following properties:

- Positivity: $\forall k \ \ r_k \geq 0, \quad t_k \geq 0$
- Monotonicity : $\forall k \ r_{k+1} \geq r_k, \quad t_{k+1} \geq t_k$
- Divergence: $\lim_{k \to \infty} r_k = +\infty, \lim_{k \to \infty} t_k = +\infty$

Example:

- $PI(x) = \sum_{i \in I_1} \frac{-1}{f_i(x)}$ ou $PI(x) = -\sum_{i \in I_1} \log\big(-f_i(x)\big)$
- $PE(x) = \sum_{i \in I_2} (f_i(x))^2$

Mixed Penalty Methods : Convergence

The next lemme is used to demonstrate the convergence of the Mixed Penalty Methods.
Let's denote by x^k the optimal solution of $Min\{M(r_k, t_k, x)\}$

Lemme 3:

for any value of k,

- $M\left(r_k, t_k, x^k\right) \geq M(r_{k+1}, t_{k+1}, x^{k+1})$
- $f(x^k) \geq f(x^{k+1})$

Theorem 3 :

Let $\{ x^k\}$ a sequence of points generated by the Mixed Penalty Methods. Any endpoint of this sequence is a solution of (P) .

Mixed Penalty Methods: Algorithm

Entrée :

- $M(r,q,x) = f(x) + \frac{1}{r}\sum_{i\in I_1}\frac{-1}{f_i(x)} + t\sum_{i\in I_2}(f_i(x))^2$ Ou $M(r,t,x) = f(x) - \frac{1}{r}\sum_{i\in I_1}\log\big(-f_i(x)\big) + t\sum_{i\in I_2}(f_i(x))^2$
- $x^0, \delta\ , r_0, t_0\ , \varepsilon_0, \alpha > 1, \beta > 1, \delta > 1\ \lim_{k\to\infty}\varepsilon_k = 0$
- $\|\nabla I(r_0, t_0, x^0)\| \leq \varepsilon_0 \quad x^0 = argmin I(r_0, t_0, y)$

- $k = 0$
- WHILE($\left\|\nabla M\left(r_k, t_k, x^k\right)\right\| > \delta$) {
  - $r_{k+1} = (r_k)^\alpha, t_{k+1} = (t_k)^\delta, \varepsilon_{k+1} = (\varepsilon_k)^\beta$
  - $Min\{M(r_{k+1}, t_{k+1}, x)\}$ {
    - $y^0 = x^k$
    - $t = 0$
    - WHILE($\|\nabla M(r_{k+1}, t_{k+1}, y^t)\| > \varepsilon_{k+1}$){
      - $\boldsymbol{d^t = -\nabla M(r_{k+1}, t_{k+1}, y^t)}$
      - **Déterminer $\boldsymbol{\alpha_t}$ tel que $\boldsymbol{\alpha_t = \underset{\alpha>0}{argmin}\{f\big(x - \alpha\nabla M(r_{k+1}, t_{k+1}, y^t)\big)\}}$**
      - $\boldsymbol{y^{t+1} = y^t - \alpha_t\nabla M(r_{k+1}, t_{k+1}, y^t)}$
      - $\boldsymbol{t = t + 1}$ }

    }
  - $x^{k+1} = y^*$
  - $k = k + 1$

}

- $x^* = x^k$ →RETURN x^* →Fin