ECN 6338 Cours 4

Résolution de systèmes d'équations non-linéaires

William McCausland

2022-02-09

Les problèmes univarié et multivarié

- Problème univarié : trouvez $x \in \mathbb{R}$ qui vérifie
$$f(x) = 0,$$
où $f : \mathbb{R} \to \mathbb{R}$.
- Problème multivarié : trouvez $x \in \mathbb{R}^n$ qui vérifie
$$f(x) = 0_n,$$
où $f : \mathbb{R}^n \to \mathbb{R}^n$.
- Problème multivarié, élément par élément : trouvez $(x_1, \ldots, x_n)$ qui vérifie
$$\begin{aligned} f^1(x_1, \ldots, x_n) &= 0 \\ &\vdots \\ f^n(x_1, \ldots, x_n) &= 0 \end{aligned}$$

La résolution de systèmes d'équations et l'optimisation

La solution x^* au problème d'optimisation

$$\max_{x \in \mathbb{R}} f(x),$$

où $f : \mathbb{R} \rightarrow \mathbb{R}$ et $f \in C^2$, est aussi la solution du système

$$\frac{\partial f(x)}{\partial x^\top} = 0.$$

Cependant, la résolution du système $g(x) = 0$, $g \in C^1$, est plus générale :

- La matrice jacobienne de g n'est pas forcément symmétrique
- La matrice jabobienne de ∇f est la matrice hessienne symmétrique de g.

Systèmes non-linéaires et le nombre de solutions

Dans le cas spécial $f(x) = Ax - b = 0$, où A est une matrice $n \times n$,

- si le rang de A est de n, il y a une solution unique;
- si le rang de A est moins grand, il n'y a aucune solution :

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

  ou il y a un nombre infini de solutions :

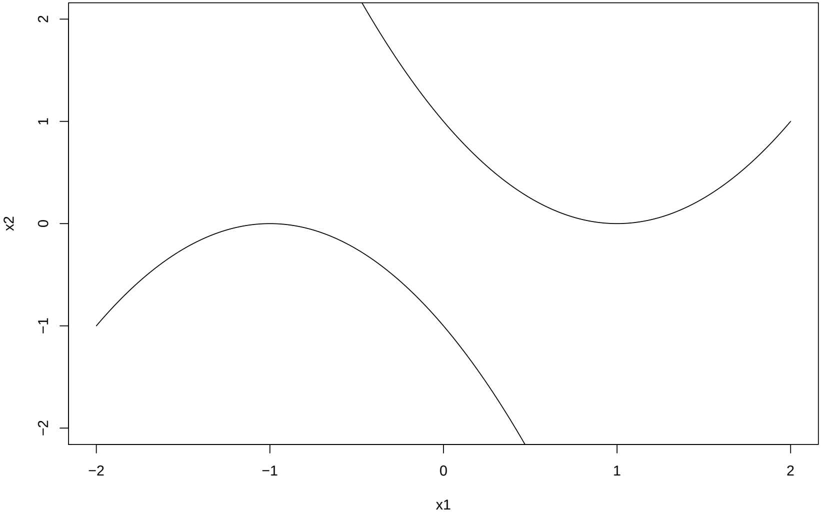
$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

Dans le cas général, il peut y avoir

- aucune solution, même pour les fonctions f^i très différentes,
- un nombre fini arbitraire de solutions,
- un nombre infini de solutions.

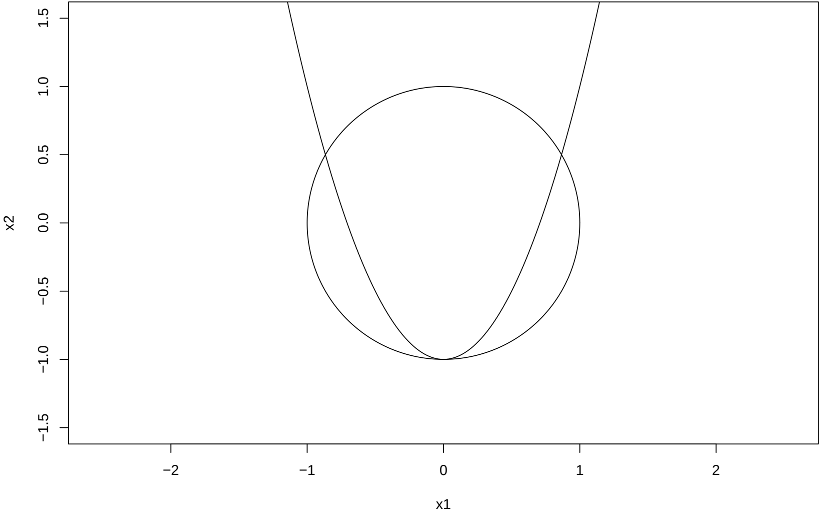
Exemple : absence d'une solution

$$f^1(x_1, x_2) = x_2 - (x_1 + 1)^2, \quad f^2(x_1, x_2) = x_2 - (x_1 - 1)^2.$$



Exemple : solutions multiples

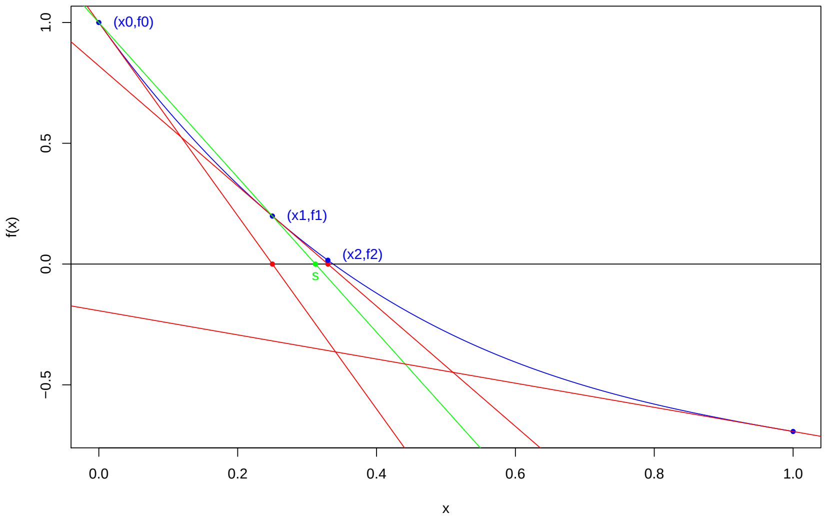
$$f^1(x_1, x_2) = x_1^2 + x_2^2 - 1, \quad f^2(x_1, x_2) = 2x_1^2 - x_2 - 1.$$



Méthodes Newton-Raphson et droite de sécante

Illustration (Newton-Raphson, droite de sécante)

```
source('root_uni.R')
```



Méthodes du type Dekker-Brent

Intrants à l'itération $k+1$: points a_k, b_k, b_{k-1} ($b{-}1 = a_0$) et valeurs $f(a_k)$, $f(b_k)$ et $f(b_{k-1})$ tels que

1. $|f(a_k)| \leq |f(b_k)|$ (point b_k, contrepoint a_k)
2. $f(a_k)f(b_k) < 0$.

À l'iteration $k+1$:

1. Calculer $m = \frac{1}{2}(a_k + b_k)$.
2. Calculer s comme fonction de a_k, b_k, $f(a_k)$, $f(b_k)$, b_{k-1}, $f(b_{k-1})$. (détails à venir)
3. Choisir entre $b_{k+1} = s$ et $b_{k+1} = m$. (détails à venir)
4. Évaluer $f(b_{k+1})$, terminer si $f(b_{k+1}) = 0$.
5. Choisir entre $a_{k+1} = a_k$ et $a_{k+1} = b_k$ tel que $f(a_{k+1})f(b_{k+1}) < 0$. (Condition 2.)
6. Si $|f(a_{k+1})| < f(b_{k+1})|$, échanger a_{k+1} et b_{k+1}. (Condition 1.)
7. Si $|a_{k+1} - b_{k+1}| < \delta$, terminer avec b_{k+1}.

Calculer s (étape 2) par interpolation linéaire (droite sécante)

$$s = b_k - \frac{b_k - b_{k-1}}{f(b_k) - f(b_{k-1})} f(b_k)$$

Notes :

1. s n'est pas une fonction de a_k.
2. Si on choisit s par interpolation linéaire, une condtion nécessaire pour choisir $b_{k+1} = s$ (étape 3) est que s se trouve entre m et b_k.

Calculer s (étape 2) par interpolation inverse quadratique

- Supposez que $f(a_k)$, $f(b_k)$ et $f(b_{k-1})$ sont distinctes
- Voici une fonction quadratique $g(y)$ qui passe par les points $(f(a_k), a_k)$, $(f(b_k), b_k)$ et $(f(b_{k-1}), b_{k-1})$:

$$\begin{aligned}
g(y) = {} & \frac{(y - f(a_k))(y - f(b_k))}{(f(b_{k-1}) - f(a_k))(f(b_{k-1} - f(b_k))} b_{k-1} \\
& + \frac{(y - f(a_k))(y - f(b_{k-1}))}{(f(b_k) - f(a_k))(f(b_k) - f(b_{k-1}))} b_k \\
& + \frac{(y - f(b_{k-1}))(y - f(b_k))}{(f(a_k) - f(b_{k-1}))(f(a_k) - f(b_k))} a_k
\end{aligned}$$

- Notez que la fonction inverse $f^{-1}(y)$ passe par les mêmes points.
- Défine $s = g(0)$, un zéro de la fonction $g^{-1}(x)$

Calculer s par interpolation inverse quadratique (cont.)

$$
\begin{aligned}
s = {} & \frac{f(a_k)f(b_k)}{(f(b_{k-1}) - f(a_k))(f(b_{k-1} - f(b_k))} b_{k-1} \\
& + \frac{f(a_k)f(b_{k-1})}{(f(b_k) - f(a_k))(f(b_k) - f(b_{k-1}))} b_k \\
& + \frac{f(b_{k-1})f(b_k)}{(f(a_k) - f(b_{k-1}))(f(a_k) - f(b_k))} a_k
\end{aligned}
$$

Notes :

1. Habituellement, c'est une amélioration, mais on peut toujours utiliser l'interpolation linéaire quand $k = 1$ où quand deux valeurs sont très près l'une à l'autre.
2. Si on choisit s par interpolation inverse quadratique, une condition nécessaire pour choisir $b_{k+1} = s$ (étape 3) est que s se trouve entre $\frac{3}{4}b_k + \frac{1}{4}a_k$ et b_k.

Choisir entre s et m (étape 3)

- $b_{k+1} = m$ est plus sécure que $b_{k+1} = s$, mais le deuxième est habituellement meilleur.
- On ajoute aux conditions nécessaires déjà mentionnées pour choisir s d'autres conditions :
  - Après un pas de bisection (pour b_k), on ajoute les conditions $|b_k - b_{k-1}| > \delta$ et $\frac{1}{2}|b_k - b_{k-1}| > |s - b_k|$.
  - Après un pas d'interpolation, on ajoute les conditions $|b_{k-1} - b_{k-2}| > \delta$ et $\frac{1}{2}|b_{k-1} - b_{k-2}| > |s - b_k|$.

Méthode de Newton

- L'expansion linéaire de Taylor autour du point actuel x^k est

$$g(x) = f(x^k) + J(x^k)(x - x^k).$$

- Si la matrice jacobienne est inversible, il y a un zéro de g à

$$x^* = x^k - J(x^k)^{-1} f(x^k).$$