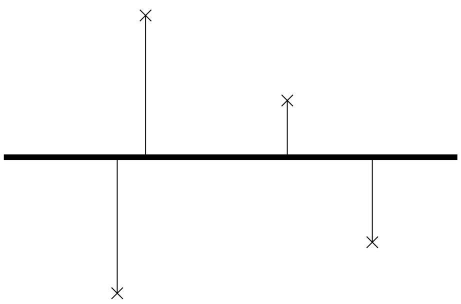
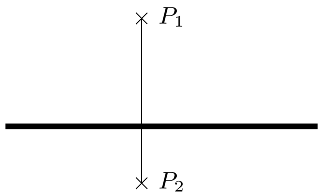
Correction TD 5 - Recherche du médian

Exercice 1 - Du pétrole et des idées



Idée simplificatrice : projeter tous les puits sur l'axe Nord-Sud.

Avec deux points seulement, la longueur totale des raccordements est minimale si l'oléoduc passe entre les puits, ou même sur l'un des puits :



Soit y_k l'ordonnée de P_k. L'altitude optimale de l'oléoduc est le médian est y_k. La chose est supposée assez évidente pour se dispenser de la prouver.

Si le nombre n de puits est pair, on a le choix entre deux médians : inférieur ou supérieur. Si n est impair, l'oléoduc passe par le puit médian.

Exercice 2 - Médian pondéré

1) On cherche à calculer le médian des x_k.

$\sum_{x_i<x_k} p_i < \frac{1}{2}$ et $\sum_{x_i>x_k} p_i \leqslant \frac{1}{2}$

$\Longrightarrow \frac{k-1}{n} < \frac{1}{2}$ et $\frac{n-k}{n} \leqslant \frac{1}{2}$

$\Longrightarrow k-1 < \frac{n}{2}$ et $n-k \leqslant \frac{n}{2}$

$\Longrightarrow \frac{n}{2} \leqslant k < \frac{n}{2}+1$

$\Longrightarrow k = \left\lfloor \frac{n}{2} \right\rfloor$ c'est le médian inférieur.

2) Algorithme simple de calcul du médian pondéré

début
  Trier les éléments : $x_{\sigma(1)} < \cdots < x_{\sigma(n)}$;
  Sommer les poids $p_{\sigma(k)}$ jusqu'à avoir : $p_{\sigma(1)} + \cdots + p_{\sigma(k+1)} \geqslant \frac{1}{2}$;
  k est alors le médian pondéré ;
fin

Complexité :
$\mathcal{O}(n \log n)$ comparaisons à cause du tri
$\mathcal{O}(n)$ additions dans la sommation

3) Algorithme en temps linéaire de calcul du médian pondéré

début

Poser $E = \{ \ (x_i, p_i) \ \}$;

si $|E| = 1$ **alors** terminé : renvoyer x_1 ;

Calculer le médian x des x_i, noter p son poids ;

Construire $E_1 = \{ \ (x_i, p_i) \ / \ x_i < x \ \}$;

Construire $E_2 = \{ \ (x_i, p_i) \ / \ x_i > x \ \}$;

Calculer $S_1 = \displaystyle\sum_{E_1} p_i$;

si $S_1 > \frac{1}{2}$ **alors**

relancer l'algorithme sur E_1 ;

sinon

relancer l'algorithme sur $E_2 \cup \{ \ (x, p + S_1 \ \}$;

fin

k est alors le médian pondéré ;

fin

Lorsque l'algorithme est relancé, il faut réajuster tous les poids pour que leur somme fasse toujours 1.
Calcul de $I(n)$, complexité d'une itération de l'algorithme, sans récursivité :
- Le calcul du médian non pondéré s'effectue en $\mathcal{O}(n)$
  (cf. cours : algorithme faisant des paquets de cinq).
- Les autres opérations s'effectuent en $\mathcal{O}(n)$ également.

$\Longrightarrow \ I(n) = \mathcal{O}(n)$

Complexité totale : $\quad T(n) = T(\frac{n}{2}) + I(n) = T(\frac{n}{2}) + \mathcal{O}(n)$

On peut alors appliquer le Master Theorem :

Notations :	$T(n) = a\,T(\frac{n}{b}) + \mathcal{O}(n^\alpha)$
Valeurs :	$a = 1 \ ; \ b = 2 \ ; \ \alpha = 1$
Cas :	$1 = a < b^\alpha = 2$
Conclusion :	$T(n) = \mathcal{O}(n^\alpha) = \mathcal{O}(n)$

$T(n) = \mathcal{O}(n)$

On a donc un algorithme linéaire pour calculer le médian pondéré.

Exercice 3 - Problème de l'emplacement du bureau de poste

On suppose les x_i triés tels que $x_1 < \cdots < x_n$.

1) Il est évident que $x \in [\, x_1 \ ; \ x_n \,]$.
On note x_m le médian pondéré des (x_i, p_i).

Soit $x \in [\, x_i \ ; \ x_{i+1} \,]$ et $\epsilon > 0$ avec $x + \epsilon \in [\, x_i \ ; \ x_{i+1} \,]$.
La situation est la suivante : $\quad x_i \leqslant x < x + \epsilon \leqslant x_{i+1}$

$$S(x) = \sum_{k=1}^{n} p_i \, |x - x_i|$$

La définition de x impose que $S(x)$ soit minimal.

$$S(x) = \sum_{x_k \, < \, x} p_k \, (x - x_k) + \sum_{x \, < \, x_k} p_k \, (x_k - x)$$

$$\begin{aligned}
S(x+\epsilon) = & \sum_{x_k \leqslant x_i} p_k\,(x+\epsilon-x_k) + \sum_{x_k > x_i} p_k\,(x_k - x - \epsilon) \\
& = S(x) + \epsilon \left(\sum_{x_k \leqslant x_i} p_k - \sum_{x_k > x_i} p_k \right) \\
& = S(x) + \epsilon \left(\sum_{x_k \leqslant x_i} p_k + \sum_{x_k \leqslant x_i} p_k - 1 \right) \qquad \text{car :} \quad \sum_{k=1}^{n} p_k = 1
\end{aligned}$$

$$S(x+\epsilon) = S(x) + \underbrace{\epsilon}_{>0} \left(2 \sum_{x_k \leqslant x_i} p_k - 1 \right)$$

$$S(x+\epsilon) < S(x) \iff \sum_{x_k < x_{i+1}} p_k < \frac{1}{2}$$

Si $x < x_{i+1} < x_m$ alors $\displaystyle\sum_{x_k < x_{i+1}} p_k < \sum_{x_k < x_m} p_k < \frac{1}{2} \implies S(x+\epsilon) < S(x)$

$\implies S(x)$ n'est pas minimal donc x n'est pas un médian pondéré.

On en déduit : $x \geqslant x_m$.

En posant $\epsilon < 0$, on montre avec un raisonnement similaire que $x \leqslant x_m$, ce qui permet de conclure : $x = x_m$. Corrolaire : l'exercice 2 n'a pas servi à rien !

2) On calcule indépendamment x_k médian sur les x, et y_l médian sur les y.
La solution au problème est le point (x_k, y_l).

Note 1 : ne pas commettre l'erreur de se limiter au calcul d'un seul médian et de déclarer solution le point (x_k, y_k). En dimension 2, le point solution n'est pas toujours confondable avec un des P_k comme c'était le cas dans la question précédente.

Note 2 : cette solution simple convient car on utilise la distance de Manhattan. Elle ne fonctionnerait pas avec la distance usuelle $d(A,B) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$.