Méthodes de descente

Marion Foare, Gaëtan Frusque, Sébastien Lérique

Objectifs.

- Compréhension des méthodes de descente, et en particulier de la **méthode de descente de gradient**
- Implémentation de la méthode en 1D et en 2D
- Compréhension du modèle de Tikhonov
- Compréhension de l'apport de la méthode par rapport aux méthodes de résolution directes
- Compréhension des limites de la méthode

Evaluation. En fin de module, vous devrez avoir rédigé un compte-rendu (max. 10 pages) sous forme de cours sur les méthodes de descente, et en particulier les méthodes de descente de gradient. Le format est semi-libre : il peut s'agir d'un document pdf, ou d'un document partagé type Google doc ou Overleaf.

Chaque semaine, un retour commenté sera fait sur votre travail, afin de corriger les erreurs, ou vous inciter à compléter certaines parties notamment.

- Si vous travaillez sur un document local, il vous est demandé de le déposer **le lundi de chaque semaine** sur le dépôt CPe-campus dédié
- S'il s'agit d'un document partagé, vous déposerez sur le dépôt un unique pdf contenant le lien vers votre document partagé.

L'évaluation de TP portera sur la compréhension des notions, évaluée en séances, l'évolution de votre compte-rendu et le compte-rendu final.

Nous rappelons que toute tentative de copie entrainera une sanction de l'ensemble des binômes concernés.

Déroulement. Ce TP est à effectuer par binôme sous Matlab. Vous trouverez sur CPe-campus une archive contenant l'ensemble des fichiers nécessaires à la réalisation de ce TP.

Ce TP fait partie intégrante de la construction de votre cours sur les méthodes de descente de gradient. Il doit vous permettre à la fois de comprendre le fonctionnement de ces méthodes, d'illustrer votre cours, mais aussi de mettre en évidence l'importance de certaines hypothèses, l'apport de ces méthodes par rapport à des méthodes de résolution classiques, et les limites de ces méthodes. Il est dès lors indispensable que vous preniez le temps d'étudier l'influence des paramètres et, le cas échéant, de vos choix d'implémentation.

Bien qu'il soit relativement guidé, n'hésitez donc pas à sortir des sentier battus, à vous poser vos propres questions et à prendre du recul sur les pistes de réflexion proposées dans ce TP.

1 Contexte

On cherche à résoudre un problème d'optimisation, défini sous la forme :

$$\hat{x} \in \underset{x \in D}{\operatorname{argmin}} \ f(x)$$

Comme évoqué en cours, l'une des principales questions liées à l'optimisation est l'implémentation d'un algorithme qui nous permette d'approcher la solution même lorsqu'aucune forme explicite n'est envisageable, autrement dit de construire une suite $\{x_n\}_n$ qui converge vers $\hat{x}$.

Bien que chaque problème d'optimisation soit unique, et qu'il n'existe pas de méthode idéale permettant de résoudre efficacement tous les problèmes d'optimisation, les **méthodes dites "de descente"** font partie des algorithmes (itératifs) les plus classiques pour approcher numériquement le minimiseur $\hat{x}$. L'idée générale repose sur le fait de rechercher, à chaque itération, x_{k+1} tel que $f(x_{k+1}) < f(x_k)$.

On se concentrera dans ce TP sur des exemples en débruitage et déconvolution 1D (signal) et 2D (image). Pour autant, l'optimisation est un domaine transversal, et permet, en fonction du choix de la fonction coût, de traiter de nombreuses applications.

2 Prise en main de la méthode

Cette section doit vous permettre de commencer à établir un plan de votre cours sur les méthodes de descente, et en particulier de la méthode de descente de gradient, que vous pourrez ensuite compléter au fur et à mesure.

Pour étudier le principe de la méthode de gradient, on se propose d'appliquer l'algorithme pour minimiser la fonction $f(x) = x^2$, i.e. résoudre le problème de minimisation 1D

$$\underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \ x^2 \tag{1}$$

Pistes de réflexion :

1. résultat attendu ?
2. implémentation de la méthode de descente de gradient
3. quel choix pour l'initialisation ?
4. quel choix du pas de descente ?
5. étude de la convergence de la méthode

3 Résolution du modèle des moindres carrés

Nous avons vu en cours que les fonctions de coût en image sont de la forme

$$f(x) = \mathcal{L}(Hx; z) + \lambda R(x)$$

où $\mathcal{L}$ est l'attache aux données (choisie en fonction du type de dégradation) et R est une régularisation (choisie en fonction du type de résultat souhaité).

Afin d'étudier l'influence de chaque terme, on se concentre dans un premier temps sur une fonction de coût sans régularisation (i.e. $f(x) = \mathcal{L}(Hx; z)$). Dans le cas d'une dégradation par un flou H et un bruit gaussien, le problème de minimisation est donc réduit au **modèle des moindres carrés**, défini par :

$$\underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \quad \|Hx - z\|_2^2 \tag{2}$$

On pourra se limiter dans un premier temps au choix $H = Id$. Une fois la méthode implémentée et validée, on pourra introduire une dégradation de type flou et/ou décimation. Pour cela, vous pouvez utiliser les fonctions `matH_1D.m` et `subsampling.m` fournies pour créer la matrice H.

Attention. On veillera à choisir un signal de taille raisonnable ($N \leq 256$), afin que la matrice H ne nécessite pas un emplacement mémoire trop important.

Pistes de réflexion :

1. solution attendue ?
2. solution analytique du problème (2)
3. résolution numérique par descente de gradient du problème (2) pour un signal 1D z de votre choix (e.g. Heaviside, polynôme, etc.)
4. comparaison des solutions obtenues par les deux méthodes (valeurs de la fonction de coût)
5. influence du pas de descente de gradient
6. influence du choix de l'initialisation
7. limites de chacune des deux approches ?

4 Résolution du modèle de Tikhonov 1D

On souhaite étudier l'influence de la régularisation R. L'un des modèles les plus connus (utilisé par exemple pour la reconstruction tomographique) en traitement d'image est le modèle de Tikhonov, défini de la manière suivante :

$$\underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \quad \|Hx - z\|_2^2 + \lambda \|\Gamma x\|_2^2 \tag{3}$$

où Γ modélise un opérateur de votre choix (identité, gradient, Laplacien...).

Pistes de réflexion :

1. solution attendue ?
2. solution analytique du problème (3)
3. résolution numérique par descente de gradient du problème (3) pour le même signal 1D que dans la partie 3
4. influence du choix de l'opérateur Γ
5. comparaison des solutions obtenues par les deux méthodes (valeurs de la fonction de coût)
6. apport de la régularisation ?
7. influence du choix de λ

5 Résolution du modèle de Tikhonov 2D

L'extension aux images est presque immédiate si l'on considère x le vectorisé de l'image. Les seuls changements à effectuer concernent les matrices H et Γ (gradient 2D, Laplacien 2D). Attention cependant à ne pas choisir des images trop grandes : on rappelle que pour une image de taille $N \times N$, Γ est de taille $N^2 \times N^2$...

Pour palier à cela, on pourra utiliser les opérateurs `D`, `L` et `H` fournis, ainsi que leurs adjoints `Dadj`, `Ladj` et `Hadj`, plutôt que les matrices.

Pistes de réflexion :

1. résolution numérique par descente de gradient du problème (3) pour une image
2. influence du choix de l'opérateur Γ
3. influence du choix de λ
4. commentaires