Jupyter Notebook : introduction et usages

Anne Cadiou

Laboratoire de Mécanique des Fluides et d'Acoustique

Ateliers et Séminaires Pour l'Informatique et le Calcul Scientifique PMCS2I - LMFA Vendredi 6 novembre 2020









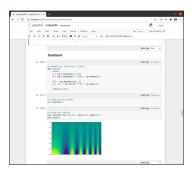




Introduction

Un Jupyter Notebook est une application web open source qui permet de créer et partager des cahiers de laboratoire, documents regroupant dans une même page du texte, du code, des équations, des graphiques, etc.





0000

Dans le cadre de développement de codes, ils peuvent aider à la documentation, la réalisation de tests. Ils sont plus particulièrement intéressants à utiliser comme feuille de synthèse, les lignes de codes étant écrites par ailleurs avec un éditeur, dans un environnement de développement. Pour ce faire, l'usage de Jupyter Lab peut être intéressant.



https://jupyter.org

Caractéristiques

Les notebooks sont des **documents JSON**, contenant une liste ordonnée de cellules (entrée/sortie) qui peuvent contenir des lignes de code, du texte (au format Markdown), des mathématiques (syntaxe proche de LATEX), des figures. Ils peuvent intégrer des fichiers mp4, etc.

Cette interface ressemble à celles de Mathematica, Maple ou SaGeMAth.

L'extension des fichiers est .ipynb

Un Jupyter Notebook peut être transformé en diapositive ou converti dans d'autres formats ouverts (HTML, présentations avec reveal.js, LATEX, PDF, ReStructuredTex, Markdown, Python) en exploitant nbconvert. Ces conversions peuvent être obtenues via la page web (Download As).

Kernel

Les kernel de Jupyter sont des environnements d'exécution des Notebooks. Pour un notebook donné, il est unique. À l'origine, il est basé sur le noyau IPython.

En général, un noyau permet l'exécution d'un langage donné.

Les noyaux ne sont pas attachés aux clients Notebooks, et peuvent être connectés aux clients en local ou par le réseau. Plusieurs clients peuvent s'y rattacher et donc partager le même environnement.

Historique

Introduction

Les Jupyter Notebooks font partie du Project Jupyter (organisation pour le développement de logiciels opensource).

Nommés en hommage aux carnets de notes de Galilée, consignant la découverte des lunes de Jupiter. Rappelle aussi les langages de base supportés par Jupyter (Julia, Python, R).

Un environnement d'exécution s'appelle un noyau (kernel).

- 2014 : Fernando Pérez crée le Project Jupyter, à partir d'IPython;
- 2015 : GitHub s'accorde avec Project Jupyter pour que le format des notebooks soit lu sur sa plateforme;
- 2018 : Première version stable de JupyterLab (interface utilisateur regroupant les notebooks, terminal, éditeur de texte, navigateur, etc.);
- 2019: Première version stable de JupyterHub (1.0), un serveur multi-utilisateurs pour les notebooks.

Actuellement, les notebooks s'interfacent avec différents environnements d'exécution dans une douzaine de langages.

Évolutions

Collaborer ou former via un notebook se popularise.

Créer son notebook et le partager peut se faire sur la plateforme Binder, http://mybinder.org.

Les notebooks permettent de travailler avec un noyau $\mathsf{C}, \, \mathsf{C}++$ ou Fortran.

En 2019, Google Drive exploite Colab (un environnement de type Jupyter Notebook) pour accéder à ses grilles.

En 2020, VS Code permet de travailler à la façon des Jupyter Notebooks.

Installation

En local sur sa machine, il est conseillé d'installer jupyter-notebook dans un environnement virtuel (virtualenv ou conda). Exemple d'installation avec conda :

Création de l'environnement

conda create --name envCondaPy3 python=3.8

Activation de l'environnement

conda activate envCondaPy3

Installation des packages pour les Notebook

conda install jupyter-notebook conda install anaconda-navigator

Ajouter le noyau dans le jupyter-notebook

conda install ipykernel
python -m ipykernel install --user --name=envCondaPy3

Démarrage du serveur

Démarrage du serveur de Jupyter Notebook

jupyter-notebook

ou

jupyter notebook

Cela ouvre en local dans le navigateur par défaut une page du type :

http://localhost:8888/tree

qui ressemble à :



Le serveur tourne, mais aucun notebook n'est encore lancé.

Le numéro du port (8888 par défaut) s'inscrémente automatiquement suivant le nombre de serveur démarrés. Il peut être spécifié au lancement du serveur :

jupyter-notebook --port=9999

Le serveur peut aussi être démarré sans lancer de navigateur :

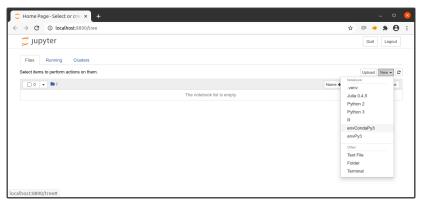
jupyter notebook --no-browser

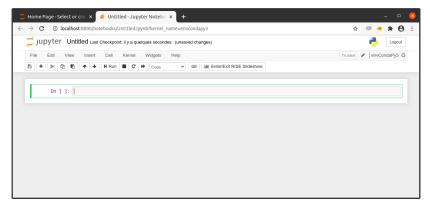
Cela peut être utile en cas de lancement du serveur sur une machine distante, à partager éventuellement avec plusieurs utilisateurs, ou sur une machine sans serveur X11. Procédure :

- Démarrer le serveur sur la machine distante;
- Créer un tunnel ssh entre la machine distante et la machine locale;
- Ouvrir un navigateur sur la machine locale en utilisant le tunnel ssh.

Création d'un notebook

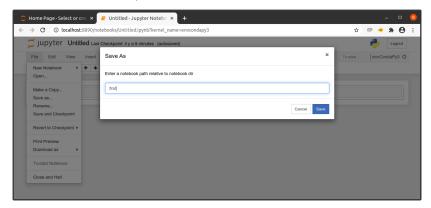
Cliquer sur le bouton New et sélectionner le noyau que vous voulez utiliser. Il doit avoir été installé sur la machine où est démarré le serveur et intégré au serveur jupyter qui a été lancé pour être dans les choix proposés.



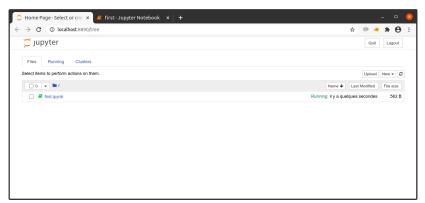


Le notebook est par défaut nommé Untitled.

Pour lui donner un nom, le sauver sous le nom voulu.

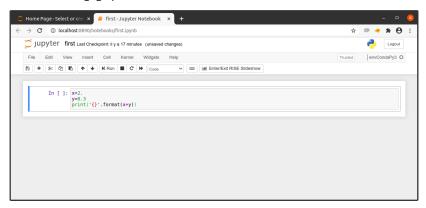


Il apparaît alors dans l'aborescence, sous ce nom :



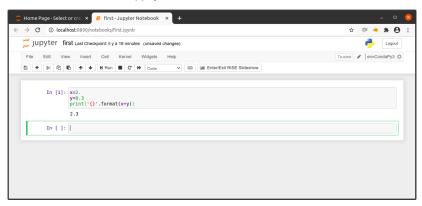
Saisir du code

Ici l'environnement envCondaPy3 est utilisé, qui est basé sur Python 3. Du code en ce langage peut donc être rentré dans les cellules du notebook.



Exécuter du code

Pour exécuter ce code, appuyer sur SHIFT+ENTER.



Les cellules sont numérotées suivant leur ordre d'exécution. Elles sont exécutées séquentiellement dans la page. Les variables définies dans une cellule sont partagées, de même que les imports. Lorsqu'on revient en arrière dans la page et ré-exécute la cellule, les variables associées sont modifiées.

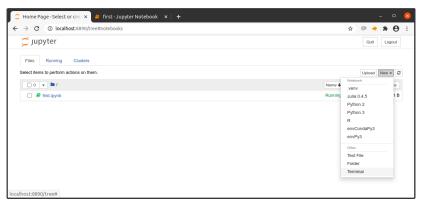
Menus

Les notebook ont des menus, proposant des opérations couramment rencontrées dans des applications et d'aures plus spécifiques.

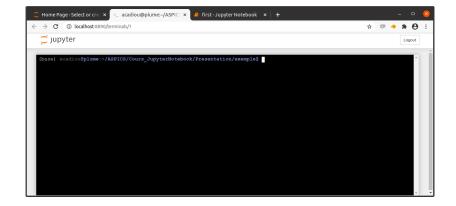
- File Manipule les fichiers. Permet de sauver et faire des reprises dans l'état d'exécution du notebook (Save anf Checkpoint). Permet de convertir en un autre format (Download as);
- Edit Manipule les cellules (couper, copier, réordonner, etc.);
- View Déplie ou cache les menus. Permet de numéroter les lignes dans les cellules. Ajoute le rôle de chaque cellule dans le mode Slide Show (diaporama avec node.js;
- Insert Ajoute une cellule avant ou après la cellule courante;
 - Cell Exécute une cellule ou un ensemble. Remet à zéro toutes les sorties ;
- Kernel Permet de stopper ou redémarrer le noyau, ou d'en changer;
- Widgets Ajoute du JavaScript;
 - Help Aide et lien vers la documentation, les raccourci clavier, la syntaxe Markdown, etc.

Démarrer un terminal

Pour utiliser des commandes bash, ajouter un kernel, trouver des fichiers, etc., le notebook permet de démarrer un terminal dans le navigateur.

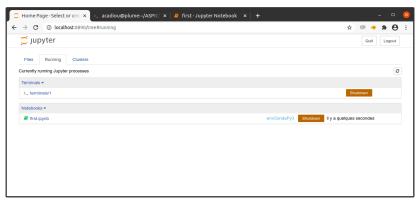


Démarrer un terminal



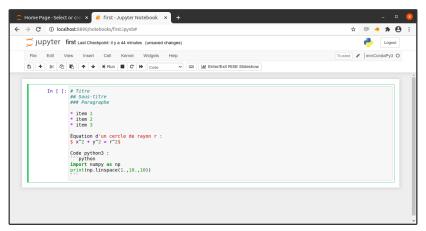
État du serveur

Dans le menu, les notifications permettent d'avoir une indication de l'état du serveur. Par notebook, cela peut aussi être consulté dans le navigateur, dans le menu Running.

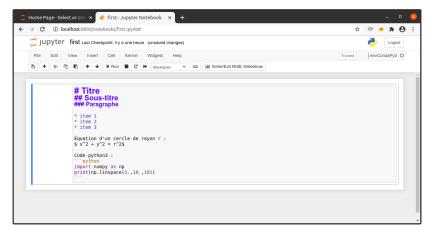


Mise en forme des cellules

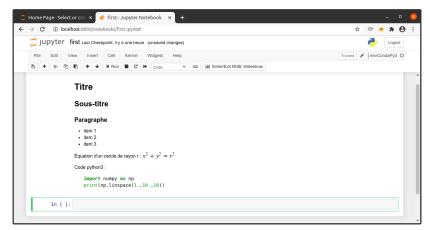
Les cellules peuvent utiliser du code ou du texte (format Markdown). Il est donc possible de mettre des sections, sous-sections, etc. d'insérer du code (qui ne doit pas être exécuté) ou des expressions mathématiques (rendu depuis LATEX).



Sélectionner le mode Markdown :



Une fois exécuté avec SHIFT+ENTER :



Exporter

Dans le menu, la section File -> Download as permet de convertir le notebookd en d'autres format. La conversion peut aussi se faire en ligne de commande avec nbconvert.

- HTML
- PLEX
- PDF
- RevealJS
- Markdown
- ReStructured Text
- Executable script

Il faut noter que nbconcert dépend de pandoc et LATEX. Ces outils doivent être présents sur la machine pour pouvoir être utilisés.

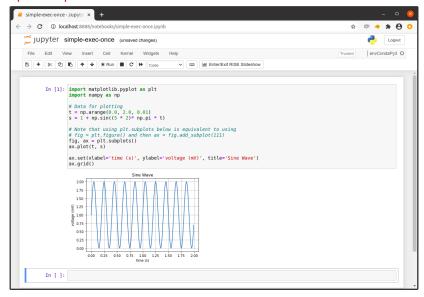
Utilisation en ligne de commande

```
jupyter-nbconvert <input notebook> --to <outputformat>

Exemple de conversion en PDF :
jupyter-nbconvert first.ipynb --to pdf

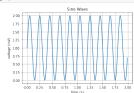
Exemple de conversion en Python :
jupyter-nbconvert first.ipynb --to python
```

Exemple d'export d'un Notebook



```
November, 6th
```

```
[1]: import matplotlib.pyplot as plt
    import numpy as mp
    t - np.arange(0.0, 2.0, 0.01)
    s - 1 + sp.sin((5 + 2)+ sp.pi + t)
    # Note that using plt.subplots below is equivalent to using
    fig, ax - plt.subplots()
    ax.plot(t, s)
    ax.set(xlabel='time (s)', ylabel='voltage (nV)', title='Sime Wave')
```

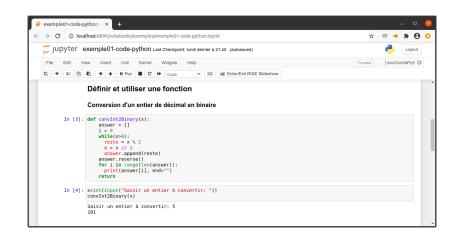


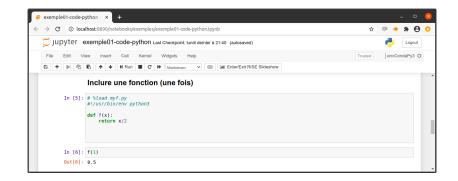
```
#!/usr/bin/env python
# coding: utf-8
# In[1]:
import matplotlib.pyplot as plt
import numpy as np
# Data for plotting
 = np.arange(0.0, 2.0, 0.01)
s = 1 + np.sin((5 * 2)* np.pi * t)
# Note that using plt.subplots below is
     equivalent to using
# fig = plt.figure() and then ax = fig.
     add_subplot(111)
fig, ax = plt.subplots()
ax.plot(t, s)
ax.set(xlabel='time (s)', ylabel='
     voltage (mV)', title='Sine Wave')
ax.grid()
# In[]:
```

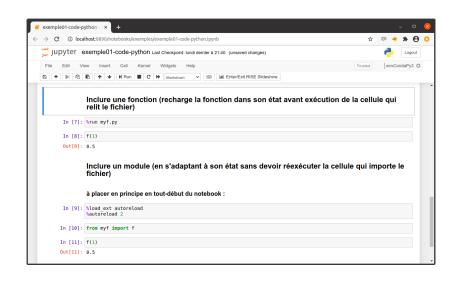
Interaction avec du code python

Exemple sur le notebook exemple01-code-python.ipynb









Remarques

Noter que l'insertion d'une fonction externe se fait : une fois pour toute, indépendamment du fichier avec la commande load :

load f.py

en s'adaptant à l'état de la fonction au moment de l'exécution de la cellule qui l'intègre au notebook, par la commande run :

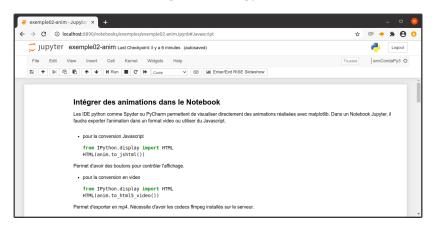
run f.py

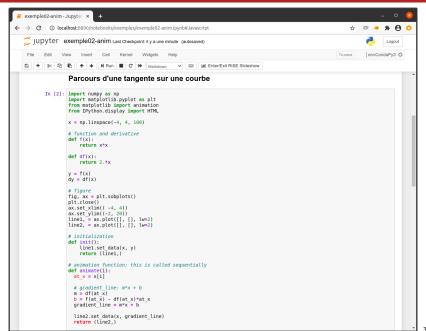
en s'adaptant à l'état de la fonction (ou du module) au moment de l'exécution de la fonction (ou des méthodes) dans leurs cellules avec l'extension autoreload

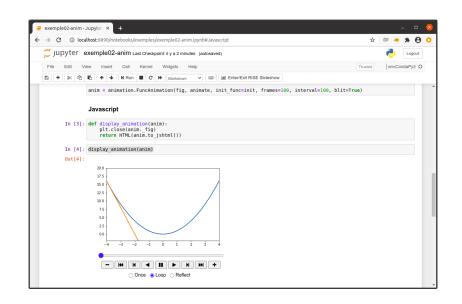
%load_ext autoreload
%autoreload 2
from myf import f
f(1)

Exemple d'animation

Exemple sur le notebook exemple02-anim.ipynb

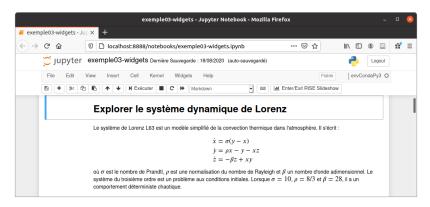


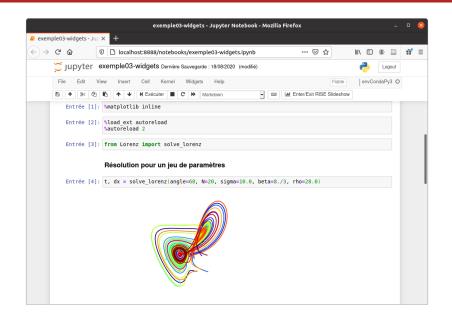


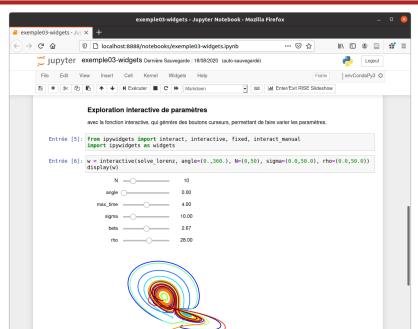


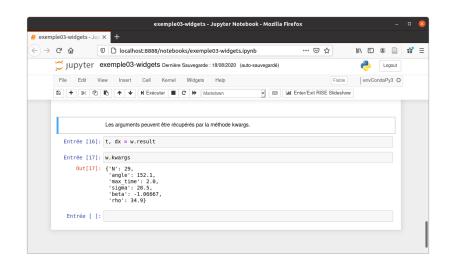
Exploration interactive de paramètres

En utilisant la fonction interactive des ipywidgets, on peut faire évoluer des paramètres avec des curseurs graphiques (et récupérer les valeurs explorées). Exemple sur le notebook exemple03-widgets.ipynb









Introduction

Un Jupyter Notebook regroupe les données, codes et résultats. Lorsque des images sont manipulées ou produites, leur versionnement peut être perçu comme un changement important par git.

Dans le cas où seuls les codes sont à versionner, ce problème peut être contourné en :

nettoyant les sorties avant tout commit;

avantage : simplicité

inconvénient : opération manuelle, ne permet pas de partager un état de reprise du

notebook

convertir en python pour le commit;

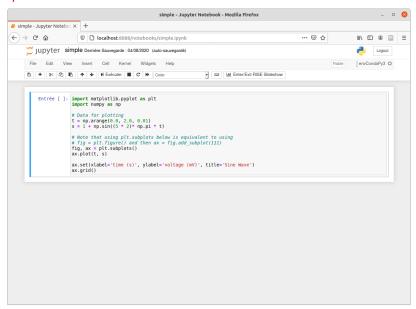
jupyter nbconvert --to="python"

avantage : code facile à versionner et partager;

inconvénient : pas intégré au notebook, donc nécessite de mettre en oeuvre la procédure

(par exemple par intégration continue)

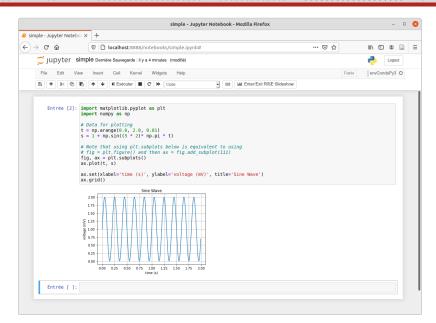
Exemple



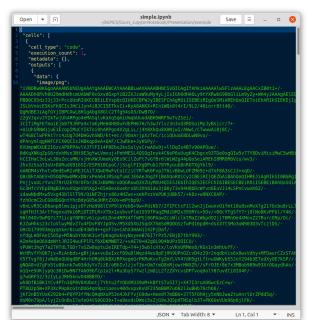
JSON

Un notebook est un fichier JSON...

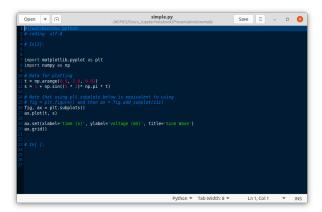
```
simple.ipynb
Open ▼ 1+1
                                                                     "cells": [
   "execution count": null.
   "metadata": {},
   "cell type": "code",
   "execution count": null.
   "metadata": {}.
   "source": []
 "metadata": {
   "display name": "envCondaPy3",
   "language": "python",
   "name": "envcondapv3"
  "language info": {
                                              JSON ▼ Tab Width: 8 ▼
                                                                        Ln 1, Col 1
                                                                                    ▼ INS
```



... dans lequel les résultats sont encapsulés.



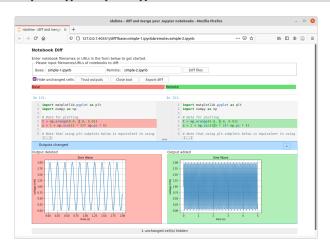
jupyter nbconvert --to python simple.ipynb



Outils utiles

https://nbdime.readthedocs.io/en/latest/ est un outil pour aider à l'analyse des différences et au versionnement.

nbdiff-web simple-1.ipynb simple-2.ipynb



Serveur distant

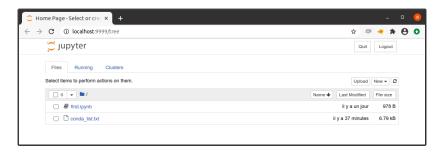
En général, les machines distantes n'ont pas de serveur graphique. Pour travailler sur ces machines, il faut donc démarrer le serveur Jupyter sans navigateur.

Par exemple, une fois l'environnement du noyau installé sur la machine distante (ici créé avec conda dans l'exemple), on voudra lancer, depuis la machine locale, un script comme :

où newton.ecl est l'alias ssh de la connexion à la machine distante. Ce script démarre le serveur distant. Pour travailler II suffit alors de définir un tunnel ssh :

```
ssh -N -f -L localhost:8889:localhost:8889 newton.ecl
```

[I 22:29:57.038 NotebookApp] Serving notebooks from local directory: /home/lmfa/ acadiou/ASPICS/exemples [I 22:29:57.038 NotebookApp] The Jupyter Notebook is running at: [I 22:29:57.038 NotebookApp] http://localhost:9999/?token=7 a23f3b493a6a80475038b1d4dc939eb0099d162ff4e0799



Pour stopper le serveur :

L'exemple est donné ici avec un environnement conda, mais un environnement virtuel avec virtualeny conviendra tout ausi bien.

JupyterLab

JupyterLab est un environnement de développement (IDE) qui s'ouvre dans le navigateur.

Installation avec conda

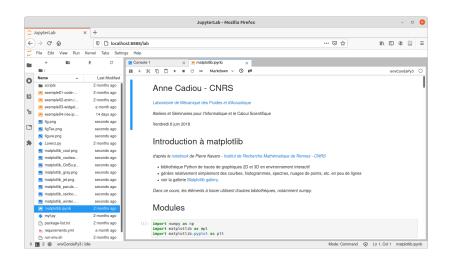
conda install -c conda-forge jupyterlab

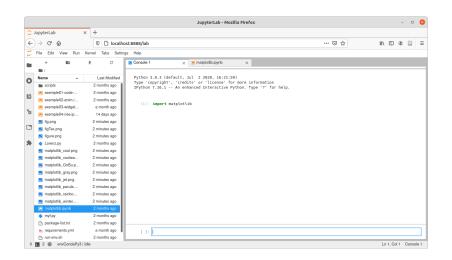
ou pip

pip install jupyterlab

JupyterLab se lance ensuite par :

jupyter lab



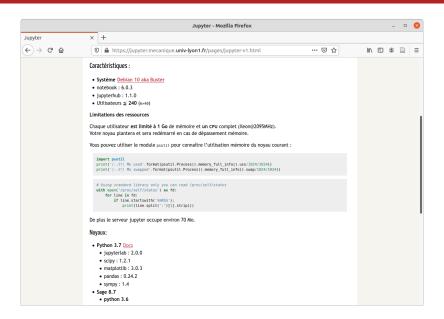


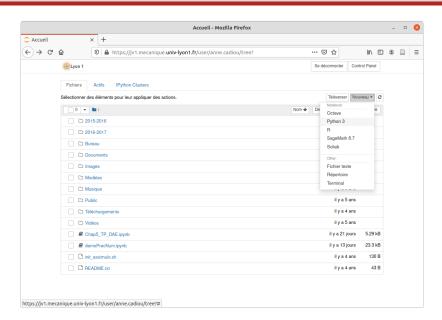
JupyterHub

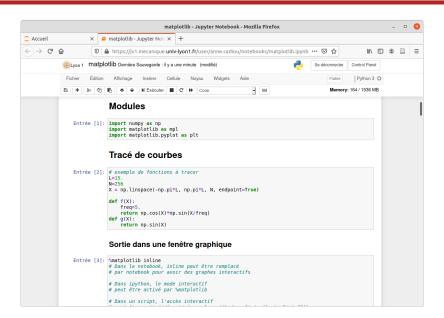
JupyterHub est un serveur multi-utilisateurs, permettant de connecter plusieurs instances de notebooks (fonctionne avec des proxy). Le Hub peut gérer des authentifications.

Exemple pour les personnels avec un compte univ-lyon1.fr





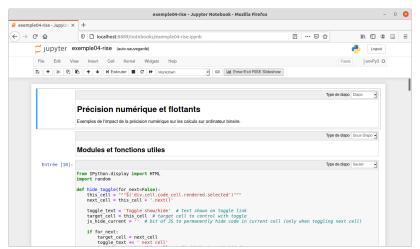


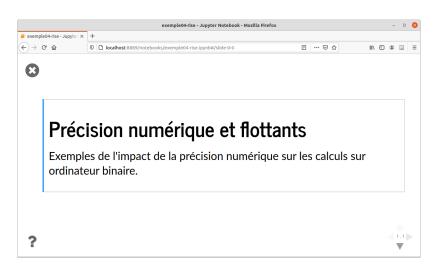


Diaporama

Les notebook peuvent être transformés en mode de diaporama. Pour cela, il suffit d'installer l'extension RISE

```
jupyter-nbextension install rise --py --sys-prefix
jupyter-nbextension enable rise --py --sys-prefix
```





Le style des pages peut être modifé via un .css

nbgrader

https://nbgrader.readthedocs.io/en/stable/

permet de préparer des feuilles d'exercices pour chaque élève, en intégrant l'évaluation par comparaison aux réponses attendues, d'intégrer des macros personnalisées en fonction de chaque élève, etc.

Références

- https://jupyter.org
- Noyau C https://github.com/brendan-rius/jupyter-c-kernel
- Noyau Fortran https://github.com/ZedThree/jupyter-fortran-kernel
- VS Code https://pbpython.com/notebook-alternative.html
- Boutons interactifs https://ipywidgets.readthedocs.io/en/stable/examples
- Versionner
 https://nextjournal.com/schmudde/how-to-version-control-jupyter
- https://nbdime.readthedocs.io/en/latest/
- https://pypi.org/project/slurm-jupyter/