# Infra

Imahn Shekhzadeh imahn.shekhzadeh@unige.ch

December 18, 2023

# **Contents**

1.	Basi	1 & Linux
	1.1.	File Download
	1.2.	for-loops
	1.3.	Argument Retrieval
	1.4.	Colored Outputs
	1.5.	String/File/Directory Operations
	1.6.	Monitoring
	1.7.	Systems Information
	1.8.	CUDA
2.	Doc	ker
	2.1.	Installation
	2.2.	Basics
	2.3.	Dockerfile
	2.4.	Docker images
	2.5.	Docker containers
		2.5.1. Basics
		2.5.2. Passing Arguments
		2.5.3. Listing & Stopping
3.	AWS	S S3 1
	3.1.	Installation & Configuration
	3.2.	AWS Credentials (Profiles)
	3.3.	Buckets
		3.3.1. Creation
		3.3.2. Listings
		3.3.3. File Copying
		3.3.4. Directory Copying
		3.3.5. Directory/File Deletion
	3.4.	Cloudpathlib
4.	Con	da 1
	4.1.	Installation of Environments
	4.2.	Export
	4.3.	Installation & Removal of Packages
	4.4.	Usage in VSCode
5.	Git	2
	5.1.	Merging
	5.2.	Merge Conflicts
	5.3.	Checking History
	5.4.	Removing a File/Folder
	5.5.	Renaming a Repository

#### Contents

	5.6. Restore File	23
	5.7. Remote URL	23
	5.8. Repo Change	
	5.9. Remote Repo Creation	23
	5.10. Pull Requests	24
	5.11. Updating Files from other Branch	24
6.	Remote Development	25
	6.1. Connection	25
	6.2. Troubleshooting	25
7.	Python	27
	7.1. Config File & JSON Files	27
	7.2. Jupyter Notebooks	27
	7.3. Map (Built-In Function)	27
	7.4. PyTorch	28
8.	Jax	31
Α.	.bashrc	33
В.	Amazing Programs, Extensions, Plugins & Packages	45
C.	Opening Programs from the CLI in Linux	47
D.	. VSCode	49
	D.1. Recommended Extensions	49
	D.2. Debugging	49
	D.3. settings.json	49
	D.4. Fix Unresolved Python Imports	49
	D.5. Opening a Duplicate Workspace	49
	D.6. settings.json	50

# Note

The commands in this document might only run through if you use the  $\verb|.bashrc|$  file provided in App. A

# 1. Bash & Linux

In Bash, [[]] over [] is preferred, since [[]] is safer and more capable within Bash scripts. Within [], where word splitting and filename expansion do occur, it is good practice to double-quote variables. But it is safe to omit the double-quotes for e.g. \$# within [[]].

#### 1.1. File Download

Downloading file from URL and allowing for redirects,

```
curl -Lo output.out https://url.com
```

When taking a GitHub link, note that you need to take the URL of the raw file.

#### 1.2. for-loops

For this directory structure,

```
infra_upd.tex
infra_upd.pdf
```

rename via

```
for file in infra_upd.*; do mv "$file" "${file/infra_upd/infra}"; done
```

What happens is a substring replacement.

## 1.3. Argument Retrieval

Retrieving all but the first argument,

```
bash_func(){
shift

echo "all provided args (except the first): $0"
}
```

Doing this  $N \geq 1$ -times,

```
test_sth(){
shift
...
shift

echo "all provided args (except the first N): $@"
}
```

## 1.4. Colored Outputs

Using colored outputs in Bash, cf. str\_diff in App. A. Personally, I use the following color scheme for the CLI,

- 1. monokai color scheme, i.e. dark gray background (#272822) with light peach color for the text (#F8F8F2).
- 2. File paths are still displayed in blue, which is suboptimal, to change the color to the better readable cyan-blue color, click on the three horizontal lines in the CLI, then on **Preferences**, then choose the currently active color, switch to the **Colors** tab, then go to **Palette**, click on the blue color & instead use the color #66D9EF

where -e stands for human readability and -s for summarizing.

## 1.5. String/File/Directory Operations

Appending line to file (-a: appending, otherwise tee overwrites output.out if existent),

```
echo "this is a line" | tee -a output.out # -a: appending, important
```

Checking whether string is empty,

```
[[ -z "$env_name" ]] && echo "The string is empty."
```

Finding out size of file/directory,

```
du -hs <path_to_file_or_dir> # du -hs file.ext

# for shorter summary (single quotation strings required)
du -hs <path_to_file_or_dir> | awk '{print $1}'
```

Unzipping a file via the CLI,

```
unzip /path/to/file.zip -d /path/to/destination
```

Opening a file and automatically scrolling to the bottom,

```
less +G /path/to/file.ext
```

Searching for files with specific extension, e.g. .ext:

```
find . -name "*.ext"
    # find . -name "*.png"
```

Creating new directory including all parent directories (-p option is safe, since if directory is already existent, no error will be outputted),

```
mkdir -p <dir>
```

Comparing the contents of two directories,

```
diff -r --color directory1 directory2 # `-r` for recursive comparison
diff -rq --color directory1 directory2 # `-q` suppresses the output of
differences and only shows which files differ
```

Ignoring files only existent in one of the directories (which treats absent files as empty),

```
diff -rq --color --unidirectional-new-file directory1 directory2
```

## 1.6. Monitoring

htop

## 1.7. Systems Information

Retrieving the number of available CPU resources,

```
echo "$(nproc)"
```

Print day and time from CLI,

```
echo "$(date +%d_%m_%y-%H_%M_%S)"

# echo "$(date +%dp%mp%y-%Hp%Mp%S)"
```

Listing all available kernels in Debian-based Linux systems,

```
dpkg --list | grep linux-image
```

Currently active kernel version,

```
uname -a
```

#### 1.8. CUDA

- When you need to find out the CUDA version installed, install nvidia-cuda-toolkit, but do NOT reboot. After its use, immediately remove this package and any package installed alongside with it!
- In case NVIDIA drivers do not allow for boot into Ubuntu, e.g. because you did not uninstall the nvidia-cuda-toolkit package,
  - 1. Boot into an older kernel version of Linux (in order to get there, do a "hard" reboot, and then go into "Advanced options for Ubuntu", and choose an older kernel version).
  - 2. Once booted into the older kernel version, I removed 'nvidia-cuda-toolkit' and rebooted.
  - 3. After a few more hard reboots and booting into the older kenel version, at some point, the newer kernel version was picked up and worker again.
  - 4. Now to fix the monitors (because dual-monitor setup didn't work), I had to open the program "Additional Drivers" and change the driver from the open-source version to an NVIDIA proprietary one.
  - 5. Then I had to install CUDA according to these instructions.
  - 6. For PyTorch to recognize the GPU, I had to reboot.

# 2. Docker

#### 2.1. Installation

- Follow this great tutorial by DigitalOcean.
- To use NVIDIA GPUs (both in PyTorch & Jax), install the NVIDIA Container Toolkit
- Once done with the installation of the NVIDIA Container Toolkit, proceed with the configuration. During the configuration, it will be necessary to restart the docker daemon, which you can achieve as follows:

```
sudo systemctl restart docker
```

#### 2.2. Basics

• Interactive start of containers:

```
d ps -a # find out ID (also docker container name)
d start -i ID
```

• Copying files from local system to docker container and vice versa; run both commands from local CLI

```
d cp file_name container_ID:/target_dir # local -> docker
d cp container_ID:/file_name dir_name # docker -> local
```

#### 2.3. Dockerfile

• When you find the command for pulling a docker image on https://hub.docker.com, e.g.

```
d pull ubuntu:jammy-20231004
```

then in the Dockerfile, just write

```
FROM ubuntu:jammy-20231004
```

When no tag is specified, by default the *latest* one will be taken. However, using the *latest* tag can potentially cause issues with reproducibility and consistency, because you might pull a different version of the image at different times without knowing it if the latest tag gets updated. For more predictable builds, it is advised to use a specific version tag.

• Note that the structure of the docker pull command is

```
d pull [OPTIONS] NAME[:TAG|@DIGEST]
```

In general, the *NAME* is in the format *repository/image*. If *repository* is not specified, Docker assumes the image is located in the default DockerHub library repository. However, many images (like PyTorch) are hosted under a specific user or organization's namespace on DockerHub, rather than the top-level library. That's why the command for the docker pull (for the latest tag) reads

```
d pull pytorch/pytorch
```

- If using a Docker image like *pytorch/pytorch:latest*, conda is already installed. In this case, the default environment is named *base*, which is a common practice in Docker images with conda unless otherwise stated.
- Copying local scripts into docker container,

```
COPY relative/path/to/script.py .
```

From the documentation:

Multiple (src) resources may be specified but the paths of files and directories will be interpreted as relative to the source of the context of the build.

It is also important to put the . at the end, since it represents the destination in the Docker image where the file should be copied. The dot . refers to the current working directory inside the Docker image, which is determined by the WORKDIR command in the Dockerfile. If WORKDIR is not set, it defaults to the root directory (/) of the image.

Also, each time the script relative/path/to/script.py changes, the Dockerfile needs to be rebuilt – however, a cached version will be used, which speeds things up.

• Copying local dirs into docker container,

```
COPY relative/path/to/dir/ .
```

• Running a Dockerfile,

```
d build -f file_name -t img_name .

d build -f file_name -t img_name:tag_name . # tag name optional, but recommended, e.g. 1.0 (no quotes required)

# d build -f file_name --no-cache -t [...] # forcing to rebuild from scratch, no cached version is used (only do if really required)
```

where img\_name will be the name of the newly created image, tag\_name the tag name and file name the name of the docker file.

Via

```
EXPOSE custom-port-number
# EXPOSE 80
```

it is possible to expose a port. Note that port exposure is related to network access. Note that even though network access might not be needed, there is still no harm in exposing a port (since an exposure of the port does not make the docker container more vulnerable).

#### 2.4. Docker images

• A Dockerfile does not necessarily need to have the name *Dockerfile*. To pass another name when building the img, do

```
d build -f custom_docker_file .
```

The . specifies the context of the build, which is the current directory in this case. I would recommend running this command from the same dir in which custom\_docker\_file is located.

Check all available Docker images via

```
d images
```

• Cleaning up dangling docker images (these are the entries with  $\langle none \rangle$  in the repository or tag name in the output of the previous algo):

```
d image prune -f
```

• Removing a Docker image – only do this when finished with using the image

```
d image rm Image_name:Tag

# d container rm <container_id> # in case some containers are using
the image
```

#### 2.5. Docker containers

#### 2.5.1. Basics

Running Docker images – without being able to utilize NVIDIA GPUs:

```
d run -it img_name # if `tag_name` was not provided
d run -it img_name:tag_name # if `tag_name` was provided during
build (recommended)
```

• Running Docker images & utilizing GPUs:

```
d run --gpus all -it img_name
d run --gpus all -it img_name:tag_name # recommended
```

• To mount a local file to the container at runtime, do

```
d run -v /absolute/path/to/script.py:/path/to/workdir/script.py --
gpus all -it img_name
d run -v /absolute/path/to/script.py:/path/to/workdir/script.py --
gpus all -it img_name:tag_name # recommended, provide `img_name`
& `tag_name`
```

The mounting expects **absolute** file paths on the side of the host machine.

• Note that you can include the bash command **pwd** to avoid having to manually pass absolute paths for the mounting

```
d run -v $(pwd)/script.py:/path/to/workdir/script.py --gpus all -it img_name:tag_name # recommended, provide `img_name` & `tag_name`
```

If you need the container to reflect changes made to the scripts on the host without rebuilding the image every time, you would use the -v flag to mount the directory. If the scripts won't change, or you don't need to reflect changes in real-time, you don't need to mount the directory, as the necessary scripts have already been copied into the image during the build process.

• It is also possible to directly mount directories:

```
d run -v $(pwd)/dir_path:/path/to/workdir --gpus all -it img_name:
    tag_name
```

Note that the specified directory from the host is mounted into the container at the specified mount point. If there are any existing files or directories in the container at the mount point, they become obscured by the mount.

- In several cases it can be useful to remove the docker container right after execution: When you...
  - ... are running many short-lived containers, like during development or testing,
  - ... want to avoid manual cleanup of stopped containers later on,
  - $\circ$  ... are running containers for one-off tasks that do not need to persist any state after they are finished.

In this case,

```
d run --rm -v $(pwd)/dir_path:/path/to/workdir --gpus all -it img_name:tag_name
```

• It is also possible to mount two separate host directories to two separate directories within the container,

```
d run --rm -v $(pwd)/dir_path1:/path/to/workdir1 -v $(pwd)/dir_path2 :/path/to/workdir2 --gpus all -it img_name:tag_name
```

This will not cause any overwriting as each -v flag creates a unique mount point inside the container.

• Finding out the python version of the Docker image

```
d run -it --rm img_name:tag_name python3 --version
```

This command will immediately remove the container after execution.

#### 2.5.2. Passing Arguments

It is possible to pass arguments when running a docker container.

1. Assuming you have a bash script run\_scripts.sh, in which a Python script, e.g.

```
#!/bin/sh
isort /app/scripts/*.py
black /app/scripts/*.py

python3 -B /app/scripts/test_script.py
python3 -B /app/scripts/test_anil.py
```

Modify this bash script s.t. any arguments passed to the CLI when running the docker container are picked up,

```
python3 -B /app/scripts/test_anil.py "$@"

# python3 -B /app/scripts/test_script.py "$@" # alternative
```

- 2. Rebuild (!) the docker image.
- 3. Now run the docker container as follows:

```
d run --rm -v $(pwd)/dir_path:/path/to/workdir --gpus all -it
img_name:tag_name arg1 arg2

# d run --rm -v $(pwd)/dir_path:/path/to/workdir --gpus all -it
img_name:tag_name --n_ways 1 --k_shots 1 # example
```

#### 2.5.3. Listing & Stopping

• Listing all running containers,

```
1 d ps
```

Listing only the container ID (of all running containers),

```
d ps -q
```

• Stopping a running container,

```
d stop container-ID
```

Stopping a running container and removing it,

```
d stop container-ID && d rm container-ID
```

# 3. AWS S3

## 3.1. Installation & Configuration

- 1. Installation instructions
- 2. The CLI will display the path under which the *aws* package was installed, but it might be sufficient to simply run

```
aws
```

Double check by running

```
which aws
```

3. After installation, configuration is necessary. For this run

```
aws configure
```

You can leave these fields empty:

```
Default region name [None]:
Default output format [None]:
```

A configuration file will be saved under

```
\sim/.aws/credentials
```

4. In the case you are a member of UNIGE, you can obtain the AWS access key ID and the secret access key as follows:

```
echo -n "$user_name" | base64 # the `-n` is important in this
context
echo -n "$passwd" | md5sum
```

where \$user\_name and \$passwd need to be provided

Otherwise, you need login to the AWS Management Console.

5. To test the configuration was successful, do this:

```
aws s3 ls --endpoint-url https://your-custom-s3-endpoint.com
```

where you replace the endpoint-url https://your-custom-s3-endpoint.com with yours.

## 3.2. AWS Credentials (Profiles)

- It is possible to use several profiles in the file  $\sim$ /.aws/credentials.
- For example,

```
[default]
aws_access_key_id = YOUR_DEFAULT_ACCESS_KEY
aws_secret_access_key = YOUR_DEFAULT_SECRET_KEY

[profile1]
aws_access_key_id = ANOTHER_ACCESS_KEY_ID
aws_secret_access_key = ANOTHER_SECRET_ACCESS_KEY

[profile2]
aws_access_key_id = YET_ANOTHER_ACCESS_KEY_ID
aws_access_key_id = YET_ANOTHER_SECRET_ACCESS_KEY
```

Using specific profile when running aws cli commands via --profile option in the command:

```
aws s3 --profile profile1 [...]
# aws s3 --profile default [...]
```

#### 3.3. Buckets

One can have several buckets.

#### 3.3.1. Creation

• Creating a new bucket,

```
aws s3api create-bucket --bucket custom-bucket-name --endpoint-url https://custom-s3-endpoint.com --profile default
```

#### 3.3.2. Listings

• Directly "folder" contents of an s3 bucket,

```
aws s3 ls s3://custom-bucket-name --recursive --endpoint-url https
    ://custom-s3-endpoint.com --profile default # `--recursive`
    optional
```

• Showing file contents,

```
aws s3 ls s3://custom-bucket-name/prefix/ --recursive --endpoint-url https://custom-s3-endpoint.com --profile default # `--recursive optional
```

Note that the / at the end of the prefix ("folder") is necessary.

#### 3.3.3. File Copying

• Local machine  $\longrightarrow$  S3:

```
aws s3 cp path/to/custom_file.ext s3://custom-bucket-name/path/to/
    custom_file.ext --endpoint-url https://custom-s3-endpoint.com --
    profile default
```

• S3  $\longrightarrow$  local machine:

```
aws s3 cp s3://custom-bucket-name/path/to/s3_file.ext custom/
    destination --endpoint-url https://custom-s3-endpoint.com --
    profile default
```

#### 3.3.4. Directory Copying

• Local machine  $\longrightarrow$  S3:

```
aws s3 sync path/to/dir s3://custom-bucket-name/path/to --endpoint-url
https://custom-s3-endpoint.com --profile default
```

#### 3.3.5. Directory/File Deletion

• Deleting a folder (which is essentially a prefix in S3) and its contents in an S3 bucket,

```
aws s3 rm s3://your-bucket-name/path-to-your-folder --recursive -- endpoint-url https://custom-s3-endpoint.com --profile default
```

• Deleting a file,

```
aws s3 rm s3://your-bucket-name/path-to-your-file.out --recursive --
endpoint-url https://custom-s3-endpoint.com --profile default
```

#### 3.4. Cloudpathlib

• When you use the cloudpathlib module, and you want to specify a profile, do this:

```
from cloudpathlib import S3Path, S3Client

# Create an S3 client with a specific AWS profile
s3_client = S3Client(
aws_access_key_id=aws_access_key_id,
aws_secret_access_key=aws_secret_access_key,
endpoint_url=endpoint_url,
profile_name="profile1", # specify profile here
)

# Make `client` default:
client.set_as_default_client()
```

# 4. Conda

• Retrieving information about currently activated conda environment,

```
conda info
```

• Listing all installed environments,

```
conda env list
```

#### 4.1. Installation of Environments

• Installing conda with specific python version,

```
# only `myenv` needs to be specified (quotation marks necessary)
env_name="myenv" && conda create -n "$env_name" python=3.11.3 -y &&
conda activate "$env_name"
```

As of Oct 16, I wouldn't recommend installing python 3.12.0 yet — I got a lot of unmet dependency problems when trying to install torch 2.1 with NVIDIA Cuda version 11.8 afterwards.

• Installation of conda environment from bash file:

```
conda deactivate # go into base environment
source conda/filename.sh
touch .env
```

• Completely remove conda environment,

```
conda deactivate && conda remove -n custom-env-name --all -y
```

#### 4.2. Export

• Exporting an .yml-file to share with others for reproducibility,

```
conda env export > environment.yml
```

• Line "Prefix:" at end of .yml file can be safely deleted, for details cf. here

## 4.3. Installation & Removal of Packages

• Installation of packages from pyproject.toml file,

```
pip install -e .
```

If there is not enough free space, do

```
TMPDIR=[...] pip install -e .
```

where TMPDIR needs to exist.

• Installing specific conda package version,

```
conda install -c conda-forge custom-pkg-name -y
# conda install -c conda-forge cloudpathlib=0.15.1 -y
```

• Removing list of packages from conda environment,

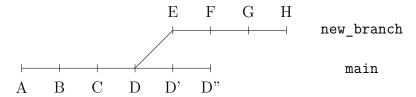
```
conda remove -n custom-env-name pkg1 pkg2 ... pkgN -y
# conda remove -n google_jax matplotlib -y
```

## 4.4. Usage in VSCode

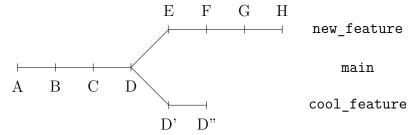
• Selecting a conda environment in VSCode, do Ctrl + Shift + P and type Python: select interpreter.

# 5. Git

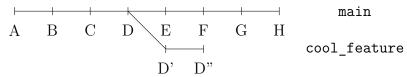
## 5.1. Merging



When merging main and new\_branch, the commits D and D'' will either be auto-merged or there will be a merge conflict. For this commit history,



merging main with new\_feature would result in a fast-forward,



Merging new\_feature into main can be done via

```
git switch main && git merge new_feature
```

Running dry merges to proactively check for conflicts if a merge was performed,

```
git merge --no-commit --no-ff branch-name && git merge --abort
```

For a fast-forward only,

```
git merge branch_to_be_merged --ff-only
```

## 5.2. Merge Conflicts

1. Resolving a merge conflict,

```
git mergetool
```

2. Confirm with Enter that you want to use vimdiff as default editing tool. vimdiff display will resemble the following structure:

```
| LOCAL | BASE | REMOTE |
| MERGED
```

If file did not already exist in BASE, then we need this view:

```
| LOCAL | MERGED | REMOTE |
```

LOCAL – Current branch

BASE – Common ancestor (how did the file look like before both changes?)

REMOTE – File that I am merging into the current branch

MERGED – Merge result

3. It is probably easiest to take the merged view and edit it directly. In the vim editor, an entire line can be deleted by pressing D (no control before!). If I instead wanted the changes from either LOCAL, BASE or REMOTE, you have to do one of these,

```
:diffg LO
:diffg BA
:diffg LO
```

Of course, the merged view can also be edited directly.

4. Type

```
:wqa
```

into vim. Afterwards, do not forget to commit and push. And if you want, do

```
git clean -f
```

## 5.3. Checking History

• Viewing the history of commits,

```
git log
```

• Viewing a specific file,

```
git show <commit-hash>:<file-name>
# git show 123abc:example.txt
```

## 5.4. Removing a File/Folder

• To remove a file/folder that is already tracked, adding it to .gitignore won't remove it (though this also needs to happen). For this, do:

```
git rm --cached <file>
git rm -r --cached <folder>
```

• Adding the file/folder to *.gitignore* is still a good idea, though, since the file/dir won't be removed locally with the commands.

## 5.5. Renaming a Repository

After having renamed the repository remotely, go to the locally cloned version of the repository and do

```
git remote set-url origin https://github.com/username/new-repo-name.git
```

double-check via

```
git remote -v
```

which lists the remote names and their URLs. No force push or the alike is necessary for the changes to take place.

#### 5.6. Restore File

Resetting specific file to state of previous commit,

```
git restore --source=<commit-hash> <file-path> && git push
git restore --source=HEAD README.md && git push
```

#### 5.7. Remote URL

Obtaining the remote URL,

```
git remote get-url origin
git remote get-url origin | sed 's/\.git$//' # optional: trim output
```

## 5.8. Repo Change

Moving all files from branch-to-move to branch-to-merge-into and preserving the commit history — do all of this while in the old repo,

```
git remote -v # check existing remotes
git remote add <target> https://target-repo-url.git # add new remote
# git remote add new-remote url
git push target branch-to-move:branch-to-merge-into
```

Do all of this in the old repo. If issue emerges during the last step, reclone the new repo and check whether this solves the issue.

#### 5.9. Remote Repo Creation

If all files to be added to the repository are in \$folder\_name,

```
folder_name="[...]" && mkdir $folder_name && cd $folder_name && git init && lpush "this is the commit msg."
```

where the command lpush is defined in App. A. Make sure there is a README.md file in the folder \$folder name.

Then install GitHub CLI and do

```
gh repo create <repository-name> --public # --private
```

Then commit and push.

## 5.10. Pull Requests

```
gh pr create --base main --head "$bname" --title "Pin isort & black versions" --body "This pull request fixes the issue that worklfows fail because of different isort/black versions used in the workflows & specified in the \`pyproject.toml\` file."
```

Note that **bname** is a bash function defined in App. A. Escaping the \ is necessary, since in shell commands, backticks (`) are used to execute commands and substitute their output into the command line.

## 5.11. Updating Files from other Branch

When working on branch\_my, it is possible to incorporate changes from another branch branch\_x,

```
git switch <branch_x>
git pull origin <branch_x>
git switch <branch_my>
git rebase -i <branch_x>
git push origin <branch_my> --force
```

Save the interactive view via :wq and make a force push to branch\_my.

# 6. Remote Development

#### 6.1. Connection

- 1. When connecting two machines remotely, install this extension on local machine (also directly in VSCode possible),
- 2. open VSCode on local machine,
- 3. press F1-button, choose "Remote-SSH: Connect to Host..." and type for the SSH host (optionally save it in the SSH config file) the same as in Algo. (B),
- 4. enter the passwd for the remote SSH host.

## 6.2. Troubleshooting

If you find you are getting a permission error for saving a file on the remote machine (in VSCode when doing the local coding), try

sudo chown custom-username path/to/custom/script.ext

custom-username here refers to the username on the remote machine. If the remote connection hung up,

fusermount -zu /path/to/dir

# 7. Python

#### 7.1. Config File & JSON Files

• When using argparse in combination with a JSON configuration file, the JSON keys need to match the long option names specified in parser.add\_argument() method calls. The argparse module itself does not automatically recognize abbreviated forms from a JSON file.

## 7.2. Jupyter Notebooks

Converting jupyter notebooks into PDFs,

```
for nb in /path/one/Notebook1.ipynb /path/two/Notebook2.ipynb [...]

do
jupyter nbconvert --to pdf "$nb"
done
```

Wildcarding notation would also work,

## 7.3. Map (Built-In Function)

• Function signature:

```
map(function, iterable, *iterables)
```

Description provided in the documentation:

Return an iterator that applies function to every item of iterable, yielding the results. If additional iterables arguments are passed, function must take that many arguments and is applied to the items from all iterables in parallel. With multiple iterables, the iterator stops when the shortest iterable is exhausted. For cases where the function inputs are already arranged into argument tuples, see *itertools.starmap()*.

• Example usage: Natively multiplying Python lists elementwise,

```
from typing import List

def multiply(x: List, y: List):
    return x * y

list_one = [i for i in range(1000)]
    list_two = [j for j in range(1000, 2000)]
    result = list(map(multiply, list_one, list_two)) # `map` is a built-in function, do not use `(list_one, list_two)` in this case
```

• Example usage: Converting NumPy arrays into PyTorch tensors,

```
a = np.array([1, 2, 3, 4])
tensor_list = list(map(torch.from_numpy, (a,))) # list containing
tensor, use of additional brackets necessary
```

• Example usage: Converting NumPy arrays into PyTorch tensors,

```
a = np.array([1, 2, 3, 4])
b = np.array([5, 6, 7, 8])
a, b = map(torch.from_numpy, (a, b)) # tuple unpacking
```

## 7.4. PyTorch

#### **Leaf Tensors**

- If requires\_grad=False, then the tensor will be leaf by convention. If requires\_grad=True, then the tensor will be leaf if it was created directly by the user and is **not** the result of an operation, e.g. .to(device) when the tensor is on cpu and device="cuda:0".
- However, by definition, leaf tensors themselves do not have a **gradient function** .grad\_fn because they are not the result of a differentiable operation applied to other tensors, i.e. grad\_fn on such tensors will return None. The gradient function in neural network libraries like PyTorch or TensorFlow is associated with tensors that are outputs of differentiable operations.
- The .grad attribute on leaf tensors that require gradients, i.e. those for which requires\_grad=True, stores the gradient computed during backpropagation. (For leaf tensors that have requires\_grad=False, calling the .grad attribute outputs None.) Note that for non-leaf tensors, calling .grad results in a UserWarning, since non-leaf tensors are generally intermediate results in the computation graph, and their gradients are usually not needed once the gradients of the leaf tensors have been obtained. However, there are cases where those gradients are needed, which can be enforced by setting retain grad=True on those tensors,

```
y = x**2

# retain gradients
y.retain_grad()

# backward pas
y.sum().backward()

# access gradients
y.grad # `torch.tensor([1., 1., 1.], device="cuda:0")`
```

Note that in the example of this code snippet, doing y.grad means that we access the gradient of the scalar loss function y.sum() — on which we performed .backward(). Correspondingly, doing x.grad implies the gradient of the scalar loss function y.sum() with respect to x.

• In general, it is **not** possible to perform **in-place** operations on leaf tensors for which requires\_grad=True, since PyTorch dynamically builds a computational graph during the forward pass, which is used during backpropagation to calculate the gradients. If leaf tensors that have requires\_grad=True are changed in-place, then the values used during the forward pass are changed, which will affect the gradient calculations in the backward pass. However, note that when no gradients are required for the operations, e.g. when performing parameter updates manually, one can use the context manager with torch.no\_grad(), in which case in-place operations on leaf tensors can be performed, since inside the context manager, requires\_grad=False.

#### Autograd & Backward

- The function torch.autograd.grad() computes the gradient. If the gradient of a scalar (loss function) wrt a (weight) matrix is taken, then the output will also be a matrix, where each element corresponds to the partial derivative of the scalar (loss function) wrt to the (weight) matrix element.
- torch.autograd.grad() is particularly useful if more direct control over the gradient computation is desired, in particular compared to .backward().
- Note that the default behavior of .backward() accumulates gradients in the .grad attribute of tensors,

#### 7. Python

```
y.sum().backward()
print(f"Gradients of `x` after second backward pass: {x.grad}") # `
torch.tensor([4., 4., 4.])`, notice how gradients have
accumulated
```

However, this behavior can be suppressed by simply zeroing the gradients, i.e. x.grad.zero\_() — note that x.grad returns a tensor, and <tensor>.zero\_() is a general PyTorch function that sets all elements in-place to 0.

# 8. Jax

Try to install via pip first. Only if this doesn't work use conda!

• Putting a Jax array onto a specific device,

• Dtype specification,

```
x = jnp.array([1, 2, 3], dtype=jnp.float32)
print(f"Dtype: {x.dtype}")
```

• Device inference,

```
x.device_buffer.device() # x: Jay array
```

• Making a Jax array out of a Python list or a Numpy array (do not use for tensors),

```
from jax import numpy as jnp

a = jnp.array([1., 2., 3.])
b = jnp.array(np.array([1., 2., 3.]))
```

• jit (just-in-time compilation): sets up a function with XLA (extended linear algebra): check out the NB test\_\_jit-compil.ipynb. Using jit,

```
import jax
from jax import numpy as jnp

@jax.jit
def selu(x: jnp.array, lamb: float = 1., alpha: float = 0.):
    return lamb * jnp.where(x > 0, x, alpha * (jnp.exp(x) - 1.0))
```

# A. .bashrc

```
ca() {
     local conda_out="$(conda env list | grep -E "$env_name" | head -n 1 |
        awk '{print $1}')"
     # check non-emptiness
     if [ -z "$1" ]; then
      echo "Usage: ca <env name>"
      return 1
     fi
     # check env existence
     if [ ! -z "$conda_out" ]; then
11
      conda activate "$1"
     else
      echo "Conda environment '$env_name' does not exist." # single quotes
14
         (') only for display
      return 1
     fi
    }
    # ----- CONDA -----
21
    # activate conda environment
22
    # usage: `ca custom-env-name`
    ca() {
     conda activate "$0"
    # deactivate currently activated conda environment
    cod() {
     conda deactivate
31
    # List all available conda envs:
    cel() {
34
     conda env list
35
    }
36
    # remove conda environment
    # usage: `crme ant-migrate-dev`
    crme() {
```

```
41
     # check number of passed arguments via `$#`
42
     if [[ $# -ne 1 ]]; then
      echo "NOTE: Exactly one argument needs to be provided"
45
      conda deactivate && conda remove -n "$1" --all -y
     fi
    }
49
    # alias for `conda__remove_packages`
    # usage (e.g.): `crm myenv pkg1 pkg2`
52
    crm() {
     conda_remove_packages "$0"
54
    # remove conda packages from environment
    # usage (e.g.): `conda__remove_packages myenv pkg1 pkg2`
    conda__remove_packages() {
     # define local variables first
61
     local env_name="$1"
     local conda_out="$(conda env list | grep -E "$env_name" | head -n 1 |
        awk '{print $1}')"
64
     # forget first argument (which is saved in `env name`)
     shift
     # check non-emptiness
     if [ -z "$env_name" ]; then
      echo "Usage: conda_remove_packages <env_name> [package1] [package2]
         ... [packageN]"
      return 1
71
     fi
72
73
     # check env existence
     if [ ! -z "$conda_out" ]; then
      conda remove -n "$env name" "$@" -y
      echo "Package(s) '$0' removed from environment '$env_name'"
     else
      echo "Conda environment '$env_name' does not exist." # single quotes
         (') only for display
      return 1
     fi
83
84
      ----- AWS ------
85
86
```

```
# helper function
87
     get__profile_endpoint_url() {
88
      # check if the first argument contains "https://"
      if [[ "$1" == https://* ]]; then
91
       local endpoint_url="$1"
92
93
       # if there's a second argument, it's the profile
       if [ -n "$2" ]; then
        local profile="$2"
       fi
98
      elif [ -n "$1" ]; then
99
100
       # if the first argument doesn't contain "https://", it's the profile
       local profile="$1"
      fi
103
104
       echo "$1 $2"
106
     }
108
     # define default vals and update based on provided args
     update__profile_url() {
      local endpoint_url="https://kalousis.s3.unige.ch"
111
      local profile="default"
112
113
      # update `endpoint_url` and `profile` if provided
114
      if [[ "$2" == https://* ]]; then
       read endpoint_url profile <<< $(get__profile_endpoint_url "$2" "$3")</pre>
      else
117
       read profile <<< $(get__profile_endpoint_url "$2" "$3") # for `</pre>
118
           endpoint url', default val will be taken
      fi
119
120
      echo "$endpoint_url $profile"
121
     }
123
     # listing
124
     # example usages (only bucket name provided):
     # `lal path`
126
     # `lal path default`
127
     # `lal path https://kalousis.s3.unige.ch`
128
     # `lal path https://kalousis.s3.unige.ch default`
     lal() {
      local path="$1"
      read endpoint url profile <<< $(update profile url "$2" "$3")</pre>
133
134
```

```
$(which aws) s3 ls s3://"$path" --recursive --endpoint-url "
135
         $endpoint_url" --profile "$profile"
     }
136
     # removing prefixes/files
138
     # example usages (only bucket name provided):
139
     # `larm path`
140
     # `larm path default`
141
     # `larm path https://kalousis.s3.unige.ch`
     # `larm path https://kalousis.s3.unige.ch default`
     larm() {
      local path="$1"
145
146
      read endpoint_url profile <<< $(update__profile_url "$2" "$3")</pre>
147
148
      command_output=$($(which aws) s3 rm s3://"$path" --recursive --endpoint
         -url "$endpoint_url" --profile "$profile")
      if [[ -z "$command_output" ]]; then
      # no use of `--recursive`, which shouldn't be used for single file
         deletion
       $(which aws) s3 rm s3://"$path" --endpoint-url "$endpoint_url" --
153
          profile "$profile"
      fi
156
     # ----- GIT ------
158
     # list all local and remote branches
159
     lb() {
      git branch -a
161
162
     # create remote branch
164
     # usage:
165
     # `lbc new-branch`
166
     1bc() {
      local branch_name="$1"
168
     git branch $(branch_name) && git push origin $(branch_name)
170
171
     # delete remote branch
173
     1bd() {
      local branch="$1"
175
      local exists__in_local=$(git branch --list "$branch")
177
      if [[ -z ${exists_in_local} ]]; then
178
      git push origin --delete "$branch" && git clean -f
```

```
else
180
       git branch -D "$branch"
181
182
     }
184
185
     # switch branches and create if non-existent
186
     lsw() {
187
      if git rev-parse --verify "$1" >/dev/null 2>&1; then
        git switch "$1"
      else
        git switch -c "$1" && git push -u origin $(bname)
191
      fi
192
193
194
     # cloning
195
     # example usage:
     # `lcl git@github.com:ImahnShekhzadeh/infra.git
     # `lcl git@github.com:ImahnShekhzadeh/infra.git infra`
198
     # `lcl git@github.com:ImahnShekhzadeh/infra.git infra main`
199
     lcl() {
200
      local dir_name="${2:-$(pwd)}"
201
      local branch_name="${3:-main}"
      git clone "$1" "$dir_name" && cd "$dir_name" && lsw "$branch_name"
204
     }
205
206
     # example usage: `lsta 2` or `lsta`
207
     lsta() {
208
      local stash_index=${1:-0} # Default to 0 if no argument provided
      # Check if the provided argument is an integer
      if ! [[ \$stash index =\sim ^[0-9]+\$ ]]; then
212
       echo "The provided index is not a valid integer."
213
       return 1
214
      fi
215
      # Check if the stash index exists
217
      if ! git rev-parse --verify stash@{$stash_index} >/dev/null 2>&1; then
218
       echo "No stash found at index $stash_index"
219
       return 1
220
      fi
221
222
      # If all checks pass, apply the stash
      git stash apply "stash@{$stash index}" --index
225
226
     # Forward commands to `git stash`
227
     lst() {
228
```

```
git stash "$0"
229
230
231
     # Stash files, if arguments are provided, they are ignored
     lstf() {
233
      git stash --include-untracked
234
235
236
     # https://stackoverflow.com/questions/19595067/git-add-commit-and-push-
         commands-in-one
     # https://stackoverflow.com/questions/14763608/use-conditional-in-bash-
         script-to-check-string-argument
     # if-else statements in bash: https://linuxhandbook.com/if-else-bash/
239
     # example usage: lgit "bit" "add ..."
240
     lpush() {
241
      (
243
      # use subshell to change directory to Git root and perform actions
      cd "$(git rev-parse --show-toplevel)" || exit
245
      git add . && git commit -a -m "$1" && git push origin $(bname) && llog
246
247
248
     }
251
     # https://stackoverflow.com/questions/3236871/how-to-return-a-string-
252
         value-from-a-bash-function
     bname() {
253
      branch=$(git branch --show-current)
254
      echo $branch
256
257
     lupd() {
258
      git fetch origin $(bname) && git log HEAD..origin/$(bname) --oneline
259
260
261
     lpull() {
      git pull origin $(bname)
263
264
265
     ldiff() {
266
      git status "$0" && git diff --color "$0"
267
268
     lforce() {
      git push origin $(bname) --force
271
272
273
     llog() {
274
```

```
git log
275
276
     1rm() {
      git rm -r "$0"
279
280
281
     lreb() {
282
      # Set default value to 5:
      num1=$\{1:-5\}
      git rebase -i HEAD∼$num1
286
287
     # Reset entire repo to state of `HEAD`, or reset specific file to a
288
         specific commit hash.
     lres() {
289
      if [[ $# -eq 0 ]] || [[ $# -eq 1 ]]; then
       local commit hash=${1:-HEAD}
291
       git reset --hard "$commit_hash"
292
      elif [ $# -eq 2 ]; then
       local commit_hash="$1"
294
       local file_path="$2"
295
       git restore --source="$commit_hash" "$file_path"
      else
       echo "Usage: lres [commit_hash file_path]"
298
299
300
301
     lsh(){}
302
      git show "$0"
303
304
305
     # usage:
306
     # `lmv file1 file2 file3 [...] /path/to/target_dir`
307
     # `lmv source_dir target_dir`
308
     lmv() {
309
       # convert the arguments to an array
       local args=("$0")
311
312
       # get the last element of the array
313
       local target_dir="${args[-1]}"
314
315
       if [[ ! -d "$target_dir" ]]; then
316
        mkdir -p "$target dir" || {
          echo "Target directory does not exist: $target dir" >&2
          return 1
319
        }
       fi
321
322
```

```
# shift the arguments so that last argument (`target_dir`) is dropped
323
       unset args[-1]
324
325
       # now, loop through all the remaining arguments
       for path in "${args[*]}"; do
327
         if [[ -e "$path" ]]; then
328
          if [[ -d "$path" ]]; then
            # move each item in the directory individually
330
            for item in "$path"/*; do
             git mv "$item" "$target dir"
            done
334
            # remove the now-empty source dir
335
            rmdir "$path"
336
          else
337
            # if it's a file, just move it
            git mv "$path" "$target_dir"
          fi
340
         else
341
          echo "File does not exist: $path" >&2
342
        fi
343
       done
344
     }
       ----- PROTONVPN -----
347
348
     p() {
349
      protonvpn-cli "$@"
350
351
                   ----- MISCELLANEOUS -----
353
354
     # pdflatex
355
     pd() {
356
      /usr/bin/pdflatex "$0"
357
358
     # convert input notebook to PDF
360
     jconv() {
361
      jupyter nbconvert --to pdf "$1"
362
363
364
     # `less` with ANSI escape characters
365
     less() {
      /usr/bin/less -R "$0"
367
368
369
     diff() {
370
      /usr/bin/diff --color "$@"
```

```
}
372
373
     # overload `shred` func, allow (recursive) shredding of dirs/files
374
     # multiple files/dirs can be provided, mixing allowed
     # usage (e.g.): `shred 10 <file_name>`
376
     # shred <file_name>
     # shred <dir path>
378
     # shred <file_name> <dir_path>
379
     shred() {
      # check whether first argument is a number
      if [[ "$1" =\sim ^[0-9]+$ ]]; then
383
       local iterations="$1"
384
       shift
385
      else
386
       iterations=5 # default
      fi
      # check file/dir existences
390
      for path in "$0"; do
391
       if check_existence "$path"; then
392
       # check whether passed input is directory or not
393
        if [[ -d "$path" ]]; then
         echo "Files to be shredded in $path:"
         find "$path" -type f -print0 | xargs -0 ls -ld
396
        fi
397
       else
        echo "Error occurred in check_existence for file/dir: $path"
399
        return 1
400
       fi
      done
403
      # prompt user for confirmation
404
      read -rp "Do you wish to proceed with shredding all files in $@ for
405
          $iterations iterations? (yes/no): " confirmation
406
      if [[ $confirmation = [yY] || $confirmation = [yY][eE][sS] ]]; then
       for path in "$0"; do
408
        if [[ -d "$path" ]]; then
409
         # shred all files within the directory
410
         find "$path" -type f -exec /usr/bin/shred -uz -n "$iterations" {} \;
411
         rm -rf "$path"
412
         echo "All files in '$path' have been shredded for $iterations
413
            iterations."
        elif [[ -f "$path" ]]; then
         # shred the individual file
415
         /usr/bin/shred -uz -n "$iterations" "$path"
416
         echo "File '$path' has been shredded for $iterations iterations."
417
        fi
418
```

```
done
419
      else
420
       echo "Shredding aborted."
421
      fi
     }
423
424
     # shortcut for clearing terminal output
425
     c() {
426
      clear
427
     }
428
     # shortcuts for exiting terminal
430
     q() {
431
      exit
432
     }
433
434
     e() {
435
      q
     }
437
438
     # tailscale
439
     ts() {
440
      tailscale status "$0"
     }
443
     # xournalpp (https://github.com/xournalpp/xournalpp)
444
     xopp() {
445
      xournalpp "$0"
446
447
448
     # strings comparison
     # usage (e.g.): `str_diff "blub1" "blub1"` or `str_diff blub1 blub1`
     # or `str_diff $(echo "hey") $(echo "hey")`
451
     # NOTE: exactly two arguments need to be provided
452
     str_diff() {
453
454
      # check number of passed arguments via `$#`
      if [[ $# -ne 2 ]]; then
       echo "NOTE: Exactly two arguments need to be provided"
457
       return 1 # return non-zero exit code to indicate error
458
      else
459
460
       # compare strings
461
       if [[ $1 == $2 ]]; then
        echo -e "Strings '$1' and '$2' \033[92mmatch\033[0m"
       else
464
        echo -e "Strings '$1' and '$2' do \033[91mNOT\033[0m match"
465
       fi
466
      fi
467
```

```
468
   }
469
    # ----- DOCKER -----
472
   d() {
473
    docker "$0"
474
475
476
    # ----- CHATGPT -----
478
   # https://github.com/kardolus/chatgpt-cli/tree/main
479
   gpt(){
480
    chatgpt -i
481
482
483
   export OPENAI_KEY=[...]
485
    # ----- ALWAYS EXECUTE -----
486
487
   {\tt add\_bit}
488
```

# B. Amazing Programs, Extensions,Plugins & Packages

- https://etherpad.org/
- https://github.com/charmbracelet/glow
- https://github.com/Oxacx/chatGPT-shell-cli
- https://github.com/kardolus/chatgpt-cli/tree/main
  - For setting the right model (cf. here for all available models),

```
chatgpt --set-model gpt-4-1106-preview --set-max-tokens 128000
```

- Usage:

```
chatgpt -i
```

- https://tailscale.com/download/
  - Once installation is complete, the command

```
sudo tailscale up
```

should be run to login, though this command will also display after installation in the CLI. The signing in should happen via GitHub. To be able to use Tailscale from a new device, it must be added as a device under https://login.tailscale.com/admin/machines. Once this is done, open a CLI and type

```
ssh name@ip_address # find out <name> and <ip_address> via
    tailscale console
# ssh ellie@100.xx.xxx.xx
```

NOTE that if the file already exists locally, it will be overwritten.

- For file copying (e.g. from the host machine to the currently used machine), do this

```
scp name@ip_address:/path/to/remote_file.ext /local/path # find
out <name> and <ip_address> via tailscale console
# ssh ellie@100.xx.xxx.xx
```

For directory copying,

```
scp -r name@ip_address:/path/to/remote_dir /local/path # find
out <name> and <ip_address> via tailscale console
# ssh ellie@100.xx.xxx.xx
```

#### B. Amazing Programs, Extensions, Plugins & Packages

- https://tailscale.com/kb/1080/cli/ (no separate installation necessary, only tailscale needs to be installed)
  - Finding out the IPv4 address of the currently active machine,

```
tailscale ip -4
```

- Finding out the IPv4 address of another machine connected via the Tailscale network,

```
tailscale ip -4 custom-name
# tailscale ip -4 ellie
```

- https://github.com/aws/aws-cli
- https://github.com/termcolor/termcolor
- LibreOffice dark theme,

 $\label{eq:colors} \textbf{Tools} \to \texttt{Options} \to \texttt{LibreOffice} \to \texttt{Application Colors} \to \texttt{Custom Colors} \to \texttt{General} \to \texttt{Document Background}, choose a dark color.$ 

## C. Opening Programs from the CLI in Linux

• Opening the settings from CLI,

gnome-control-center

• Opening VSCode from CLI:

code path\_to\_file/file\_name.ext

If a VSCode editor is already open, use the -n flag to open the file in a new editor:

code -n path\_to\_file/file\_name.ext

A folder can also be opened directly:

code path\_to\_dir

Listing C.1: Opening VSCode dir from CLI

• Opening LibreOffice from CLI:

libreoffice --writer path\_to\_dir/filename.odt

• Opening an image via the CLI:

eog /path/to/your/image.jpg

### D. VSCode

#### D.1. Recommended Extensions

- https://marketplace.visualstudio.com/items?itemName=ms-vscode-remote.vscode-remot
- https://marketplace.visualstudio.com/items?itemName=Gruntfuggly.todo-tree

#### D.2. Debugging

- Stepping into external code with Python debugger, tutorial here
- Creating a JSON file, here some instructions

#### D.3. settings.json

Opening the file,

- 1. press Ctrl + Shift + P to the Command Palette,
- 2. type Open User Settings (JSON) and select it to open the settings. json file.

### D.4. Fix Unresolved Python Imports

• If you run a docker container where a conda environment is installed (with packages that you do not have locally), then VSCode will show those imports as unresolved. To fix this, open the settings. json file, cf. App. D.3, and add the following setting:

Incorporating this into the settings. json file is shown in App. D.6.

• Note that if you have an SSH connection to another machine going on, e.g. in the Remote Development extension, putting the above lines into the settings.json file will not have an immediate effect, for this the SSH connection needs to be restarted.

#### D.5. Opening a Duplicate Workspace

- 1. press Ctrl + Shift + P to open the Command Palette,
- 2. then type Workspaces: Duplicate As Workspace is New Window

#### D.6. settings.json

Contents of settings.json,

```
{
      "workbench.colorTheme": "Default Dark Modern",
2
      "telemetry.telemetryLevel": "off",
      "editor.wordWrap": "wordWrapColumn",
      "editor.wordWrapColumn": 79,
      "workbench.editor.enablePreview": false,
      "gitlens.telemetry.enabled": false,
      "notebook.lineNumbers": "on",
      "explorer.confirmDragAndDrop": false,
      "window.zoomLevel": 1,
      "python.analysis.diagnosticSeverityOverrides": {
          "reportMissingImports": "none"
13
       "todo-tree.general.tags": [
14
          "BUG",
          "HACK",
          "FIXME",
          "TODO",
          "NOTE",
          "XXX",
20
          "[]",
21
          "[x]"
       ],
      "files.associations": {"*.log": "plaintext"},
      "[plaintext]": {"editor.wordWrap": "off"},
      "[shellscript]": {"editor.wordWrap": "off"},
27
      "workbench.editor.tabSizing": "shrink"
28
```