**ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ**

**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ**

**«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

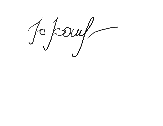
Факультет компьютерных наук

Образовательная программа «Прикладная математика и информатика»

**Отчет об исследовательском проекте**

на тему «Следы солнечной активности в температурных рядах»

(промежуточный, этап 1)

**Выполнил**:

студент группы БПМИ188 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Ю.М. Кошелева 7.02.2020

Подпись И.О. Фамилия Дата

**Принял**:

руководитель проекта \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Имя, Отчество, Фамилия

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Должность

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Место работы

Дата \_\_\_\_\_\_\_\_\_2020 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Оценка (по 10-тибалльной шкале) Подпись

**Москва 2020**

Содержание

[Введение 3](#_Toc32010913)

[Описание метода 4](#_Toc32010914)

[Проделанная работа 5](#_Toc32010915)

[Заключение 10](#_Toc32010916)

[Список источников 11](#_Toc32010917)

# Введение

В настоящее время существует множество способов и методов для анализа и прогнозирования временных рядов, с которыми люди сталкиваются постоянно. Например, современная проблема влияния солнечной активности на температуру на Земле является одной из важнейших. Мир озадачен климатическими изменениями, факторами, которые на это влияют, и прогнозированием дальнейших перемен. Для выявления зависимости между температурными изменениями и солнечной активностью используют так называемый анализ сингулярного спектра (Singular spectrum analysis).

Этот анализ базируется на преобразовании одномерного временного ряда в многомерный ряд с последующим применением к полученному многомерному временному ряду [метода главных компонент](https://ru.wikipedia.org/wiki/%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_%D0%B3%D0%BB%D0%B0%D0%B2%D0%BD%D1%8B%D1%85_%D0%BA%D0%BE%D0%BC%D0%BF%D0%BE%D0%BD%D0%B5%D0%BD%D1%82). Способ преобразования одномерного ряда представляет собой «свёртку». Для этой свёртки необходимо определить длину «окна». Чаще всего её выбор обусловлен признаками периодичности в исходных данных.

Анализ сингулярного спектра хорош тем, что подходит для решения множества задач, таких как нахождения общего тренда в ряде данных, подавление шума, выявление периодичностей, сглаживание. В связи с этим он широко применяется в климатологии, океанологии, геофизике, обработке изображений и т.д.

Выбрав тему, я поставила перед собой цель исследовать температурные ряды, характеризующие температуру на Земле, и временные ряды, характеризующие солнечную активность, на зависимость посредством анализа сингулярного спектра. Чтобы приблизиться к этой цели мне нужно выполнить следующие задачи:

1. Ознакомиться с методом SSA-разложения

2. Рассмотреть искусственно полученную функцию как сумму других

3. Рассмотреть SSA-разложение этой функции

4. Выявить зависимости между видом функции и ее SSA разложением

В первой части работы предметом моего исследования выступит анализ сингулярного спектра, а объектом соответственно его реализация на языке Python.

# Описание метода

Метод обработки данных можно разделить на три последовательных этапа.

1. Построение траекторной матрицы.

Для построения траекторной матрицы нужно установить длину окна L. Далее нужно построить матрицу размера KхL, где каждый столбец под номером *i* будет соответствовать исходным данным с индексами от *i* до *i + L – 1*.

1. Сингулярное разложение

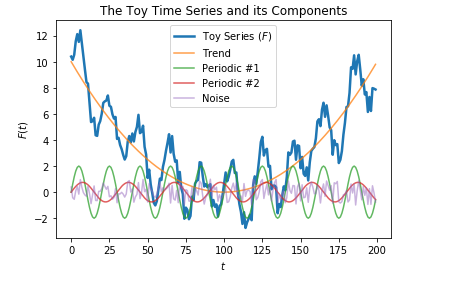
Далее к полученной траекторной матрице применяется метод сингулярного разложения. Среди полученных данных значения вектора *Sigma* будут сингулярными значениями компонент.

1. Получение компонент

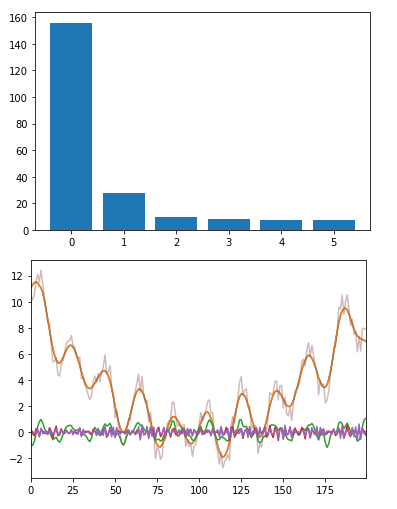
Далее к элементарным матрицам применяется усреднение по анти-диагоналям (таким образом получаем Ганкелевы матрицы). Получаем разложение исходного временного ряда данных на компоненты.

# Проделанная работа

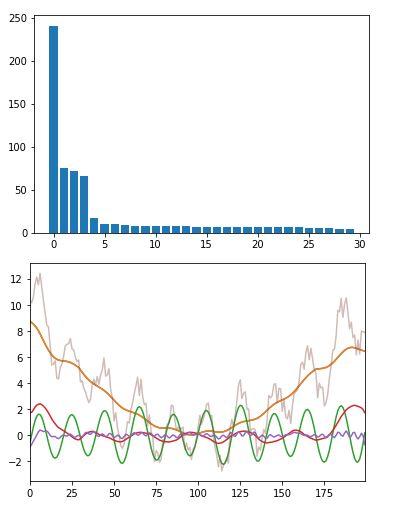
1. Изначально я построила функцию для дальнейшей работы. Эта функция складывалась из тренда(параболы), двух синусоид с разными коэффициентами и периодами и шума (значения этой компоненты генерировались рандомно). Функция стала суммой всех компонент.



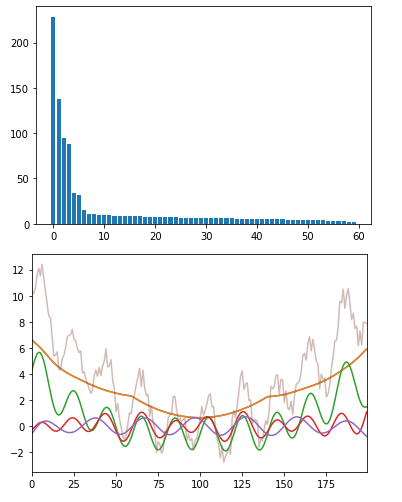
1. Далее к полученной функции несколько раз был применен SSA, но с переменной длиной окна. Соответственно, менялись и полученные компоненты. Параллельно для каждого разложения я строила график сингулярных значений. Сингулярное значение компоненты должно показывать «вклад» этой компоненты в исходную функцию.



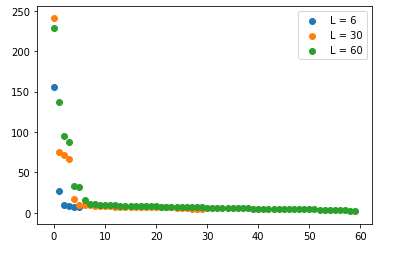
(разложение с длиной окна 6)



(разложение с длиной окна 20)



(разложение с длиной окна 60)



(График для сравнения сингулярных значений при смене длины окна)

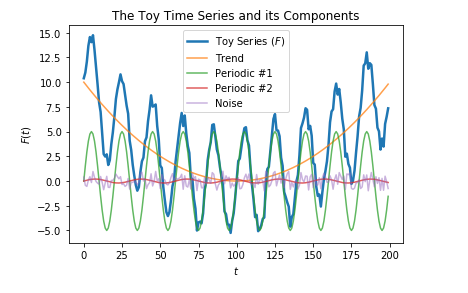
Вывод: SSA позволяет достаточно точно восстановить компоненты сигнала). Наилучший вариант разложения на компоненты достигается, когда характерные особенности функции попадают внутрь окна. В нашем примере это происходит, когда размер окна соизмерим с обоими периодами.

При длине окна равной 6 периодичности функции, которую создают синусоиды, не просматривается. В этом случае разложение имеет всего одну наиболее выделяющуюся компоненту, которая наиболее близка к самой функции, но не отображает ни тренд, ни периодичность, ни шум. При длине окна 60 на графике сингулярных значений появляется множество компонент, которые соизмеримы друг с другом, появляется множество синусоид, которые не несут в себе смысла и не относятся к изначальному разложению функции.

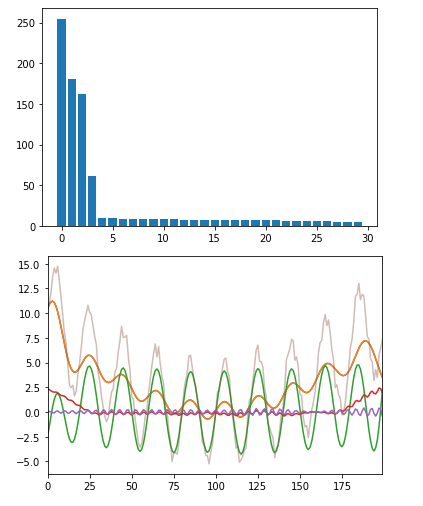
Наиболее точно восстановить исходные компоненты получается при длине окна 30. Также на графике сингулярных значений можно увидеть 4 наиболее выделяющихся компоненты. Это и есть тренд, синусоиды и шум.

Для дальнейшей работы я использовала L = 30, так как сочла эту длину окна оптимальной и наиболее точной для разложения на компоненты этой функции.

1. Далее я попробовала поменять коэффициенты синусоид для того, чтобы посмотреть, как изменятся сингулярные значения компонент при этом. Я увеличила коэффициент одной синусоиды с 2 до 5, и уменьшила у второй с 0.75 до 0.2.



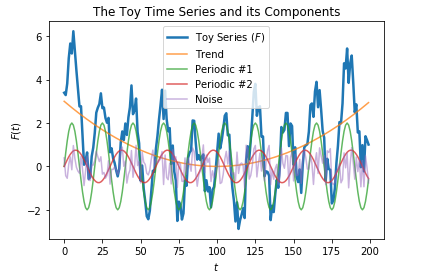
(график описанной функции и компонент, которые ее составляют)



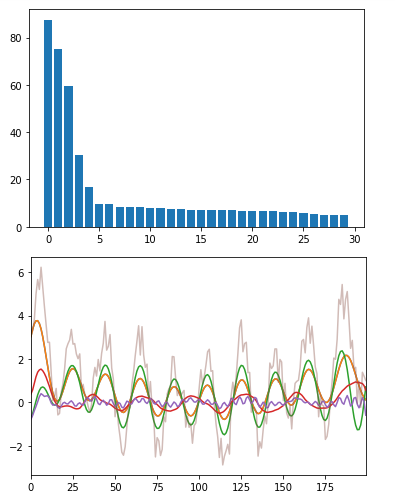
(график сингулярных значений и график компоненты после разложения)

Вывод: так как влияние одной синусоиды на функцию возросло, это отображается на графике сингулярных значений. Тренд на нем уже не выделяется так явно. Компонента тренда и компоненты синусоид имеют сравнимые сингулярные значения. Этот эффект прослеживается и на графике самой функции. На ней более заметны колебания, чем основной тренд.

1. Теперь я изменила коэффициент параболы, которая задает тренд, тем самым сделала ее более пологой.



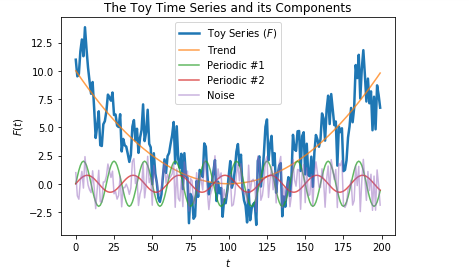
(График функции и составляющих ее компонент)

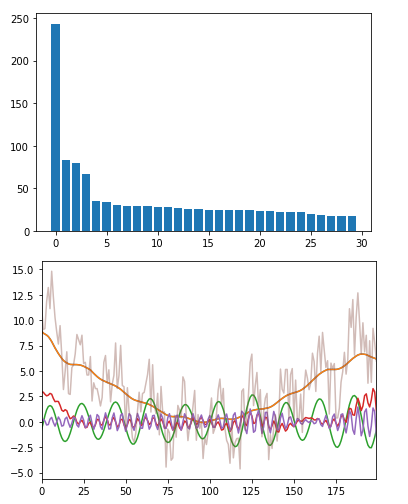


(график сингулярных значений и компонент после разложения)

Вывод: на графиках прослеживается эффект, схожий с пунктом 3. Парабола стала более пологой, соответственно на фоне горизонтальных синусоид она выделяется меньше. На графике сингулярных значений синусоиды наиболее приближены к тренду. Компоненты становятся соизмеримы друг с другом, что усложняет задачу анализу сингулярного спектра, так как выделение отдельных компонент из общей массы становится сложнее.

1. Увеличение коэффициента шума.





Вывод: шум заметно и значимо изменил вид функции, но на разложение это повлияло меньше, чем остальные преобразования. Тренд выделен четко и имеет существенное сингулярное значение, по сравнению с другими компонентами. В свою очередь шум приближен к синусоидам, но это также не мешает их выделению.

### Заключение

В ходе первых этапов работы (ознакомление с анализом сингулярного спектра) мной был изучен метод обработки временных рядов, активно применяемый на практике. Также путем построения искусственных данных, я выявила некоторые закономерности и зависимости между видом функции, подаваемой на вход SSA, и выходными данными анализа, а именно компонентами функции и их сингулярными значениями. В дальнейшем я смогу применить этот метод для поиска зависимостей между временными данными, характеризующими солнечную активность и характеризующими температуру на Земле.

# Список источников

1. Интернет-источник

<https://www.kaggle.com/jdarcy/introducing-ssa-for-time-series-decomposition> - реализация метода, пример работы на искусственной функции

1. Интернет-источник <https://ru.wikipedia.org/wiki/SSA_(%D0%BC%D0%B5%D1%82%D0%BE%D0%B4)> – статья на википедии
2. Интернет-источник

<https://github.com/aj-cloete/pssa> - репозиторий с реализацией и SSA и примерами