

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный университет имени М.В.Ломоносова»
Химический факультет

УТВЕРЖДАЮ

Декан химического факультета,

Акад. РАН, профессор



/С.Н.Калмыков/

«30» августа 2022 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Введение в MATLAB и регрессионный анализ
(Introduction to MATLAB and regression analysis)

Уровень высшего образования:

Программа подготовки научных и научно-педагогических кадров в аспирантуре

Физическая химия (104-01-00-144-хн-фмн)

Москва 2022

1. Краткая аннотация:

Название дисциплины - Введение в MATLAB и регрессионный анализ (*Introduction to MATLAB and regression analysis*)

Цель изучения дисциплины – познакомить слушателей с основами языка программирования MATLAB и сформировать у слушателей базовые знания языка программирования MATLAB, включая его применение для оценки погрешностей измерений, проверки статистических гипотез и проведения регрессионного анализа. Ключевой элемент дисциплины — это формирование компетенций, позволяющих аспирантам использовать MATLAB и регрессионный анализ как в прикладном контексте, так и в рамках научно-исследовательской деятельности. Курс состоит из трёх связанных разделов:

- основы программирования на языке MATLAB,
- оценка погрешностей и проверка статистических гипотез,
- метод наименьших квадратов.

2. Уровень высшего образования – подготовка кадров высшей квалификации

3. Научная специальность: **1.4.4** Физическая химия, область науки: 1. Естественные науки

4. Место дисциплины (модуля) в структуре Программы аспирантуры: Обязательные Дисциплины (модули) – Обязательная дисциплина по выбору.

5. *Объем дисциплины (модуля) составляет 2 зачетные единицы, всего 72 часа, из которых 58 часов составляет контактная работа аспиранта с преподавателем (18 часов занятия лекционного типа, 4 часа мероприятия текущего контроля успеваемости и промежуточной аттестации, 36 часов занятий семинарского типа), 14 часов составляет самостоятельная работа обучающегося.*

6. Входные требования для освоения дисциплины (модуля), предварительные условия.

На предыдущих уровнях высшего образования должны быть освоены общие курсы:

1. «Математический анализ»,
2. «Линейная алгебра»,
3. «Теория вероятностей»,
4. «Информатика»,
5. «Физическая химия»

7. Содержание дисциплины (модуля), структурированное по темам:

Наименование и краткое содержание разделов и тем дисциплины (модуля), форма промежуточной аттестации по дисциплине (модулю)	Всего (часы)	В том числе								
		Контактная работа (работа во взаимодействии с преподавателем), часы из них						Самостоятельная работа обучающегося, часы из них		
		Занятия лекционного типа	Занятия семинарского типа	Групповые консультации	Индивидуальные консультации	Занятия, направленные на проведение текущего контроля успеваемости, промежуточной аттестации	Всего	Выполнение домашних заданий	Подготовка к коллоквиумам	Всего
Тема 1. Основы программирования на языке MATLAB.	24	6	12				18	3	3	6
Тема 2. Оценка погрешностей и проверка статистических гипотез.	16	4	8			2	14	2		2
Тема3. Метод наименьших квадратов.	30	8	16				24	3	3	6
Промежуточная аттестация <u>зачет</u>	2					2	2			
Итого	72	18	36			4	58	8	6	14

Содержание тем:

Тема 1. Основы программирования на языке MATLAB.

Основные конструкции языка. Скрипты и функции (в файле, вложенные, анонимные). Векторизация кода. Графика и файловый ввод-вывод. Отладка программ.

Тема 2. Оценка погрешностей и проверка статистических гипотез.

Погрешности, стандартные отклонения, доверительные интервалы. Распределения и квантили. Погрешности чисел с плавающей запятой и машинный ноль. Проверка статистических гипотез с помощью критериев Стьюдента, Фишера и Пирсона.

Тема 3. Метод наименьших квадратов.

Переопределённые системы уравнений и метод наименьших квадратов. Одно- и многомерная линейная регрессия, погрешности коэффициентов и стандартная ошибка регрессии. Критерии Стьюдента и Фишера применительно к регрессии. Нелинейная регрессия, методы Ньютона, Гаусса-Ньютона, Левенберга-Марквардта. Дуальные числа и автоматическое дифференцирование. Мультиколлинеарность и гребневая регрессия. Методы глобальной оптимизации: метод отжига, генетические алгоритмы, символьная регрессия. Использование метода наименьших квадратов при решении физико-химических задач (построении фазовых диаграмм, расчёта химических равновесий, описания кинетических кривых и т.п.)

8. Образовательные технологии.

Лекции с презентациями и интерактивной демонстрацией решения задач на персональном компьютере. Семинарские занятия с решением задач на персональных компьютерах в Microsoft Excel (или Libre Office) и среде MATLAB (или GNU Octave).

9. Учебно-методические материалы для самостоятельной работы по дисциплине (модулю):

Аспирантам предоставляется программа курса, план занятий, перечень домашних заданий, лекционные материалы и примеры программ на языке MATLAB. По теме каждой лекции указывается материал в источниках из списков основной и вспомогательной литературы.

10. Ресурсное обеспечение:

Основная литература:

1. Дьяконов В.П. MATLAB. Полный самоучитель. ДМК Пресс, 2012.
2. Ю. Н. Тюрин, А. А. Макаров, «Анализ данных на компьютере». М., «Форум», 2008.
3. Д. Химмельблау, «Анализ процессов статистическими методами». М., «Мир», 1973.
4. «Справочник по прикладной статистике». Под ред. Э. Ллойда и У. Ледермана. М., «Финансы и статистика», 1989.
5. К. Дёрффель, «Статистика в аналитической химии». М., «Мир», 1998.
6. Дж. Дэннис мл., Р. Шнабель. Численные методы безусловной оптимизации и решения нелинейных уравнений. М., «Мир», 1988

Дополнительная литература:

1. F.A. Graybill, H.K. Iyer. Regression analysis. Concepts and applications. Duxbury Pr., 1994.
2. Heij C., deBoer, P., Franses, P. H., Kloek T., vanDijk. Econometric methods with applications in business and economics. Oxford University Press. 2004.
3. А.В. Гармаш, Н.В. Сорокина «Метрологические основы аналитической химии».

Перечень информационных технологий:

- Программное обеспечение для расчётов: MATLABили GNU Octave, Microsoft Excelили Libre Office. Возможно прохождение курса с использованием исключительно свободного (бесплатного) программного обеспечения, в т.ч. операционной системы.
- Справочные системы: NISTChemistryWebbook (<http://webbook.nist.gov/chemistry/>), БД «Термические константы веществ» (<http://www.chem.msu.ru/cgi-bin/tkv.pl?show=welcome.html/welcome.html>)

Описание материально-технической базы:

- Занятия проводятся в компьютерном классе

11. Язык преподавания – русский

12. Преподаватели:

канд. хим. наук, старший научный сотрудник

Восков Алексей Леонидович, alvoskov@gmail.com, 8-495-939-22-80

**Фонды оценочных средств,
необходимые для оценки результатов обучения**

Образцы домашних заданий:

- Построить график функции $f(x,y) = uexp(-x^2-y^2)$ в виде поверхности.
- Рассчитать объём полусферы с $r=2$ с помощью метода Монте-Карло, используя векторизацию кода.
- Построить изображение множества Мандельброта, используя векторизацию кода.
- Одномерная линейная регрессия.
- Многомерная линейная регрессия.
- Нелинейная регрессия с линеаризацией (описание экспериментальных данных по теплоёмкости и химической кинетике)
- Нелинейная регрессия с использованием метода Левенберга-Марквардта (оптимизация параметров термодинамической модели)
- Линейная гребневая регрессия.
- Найти глобальный минимум двумерной функции Растригина с помощью генетического алгоритма.
- Найти глобальный минимум двумерной функции Растригина с помощью метода отжига.
- Рассчитать кривую титрования H_3PO_4 , рассматривая диссоциацию одновременно по всем ступеням и полагая коэффициенты активности равными единице.
- Построить p -х и T -х диаграммы системы «ацетон-хлороформ», найдя нужные параметры стабильности в справочнике и оптимизировав параметр модели регулярных растворов, используя экспериментальные данные.

Примеры вопросов к семинарам:

- Погрешности: абсолютная, относительная, случайная, систематическая. Правила округления. Косвенные измерения.
- Дисперсия и стандартное отклонение. Квантили и доверительные интервалы.
- Проверка статистических гипотез. Критерии Стьюдента, Фишера, Пирсона. Проверка распределения на нормальность.
- Числа с плавающей запятой и машинный ноль. Оценка погрешности формул численного дифференцирования.
- Линейная регрессия. Коэффициенты регрессии, их доверительных интервалов, стандартной ошибки регрессии. Критерий Стьюдента и Фишера.
- Нелинейная регрессия: решение путём линеаризации. Влияние линеаризации на значение параметров.
- Нелинейная регрессия: численные методы (Ньютона, Гаусса-Ньютона и Левенберга-Марквардта).
- Матрица Якоби в нелинейной регрессии и её роль в оценке погрешностей. Способы её расчёта: аналитическое, численное и автоматическое дифференцирование.
- Регуляризация по Тихонову и гребневая регрессия.
- Методы глобальной оптимизации. Метод отжига и генетические алгоритмы. Символьная регрессия.

- Использование численных методов для расчёта фазовых и химических равновесий.
- Оптимизация параметров физико-химических моделей по экспериментальным данным

Вопросы для промежуточной аттестации – зачета:

A1. Теоретические вопросы по погрешностям

Абсолютная и относительная погрешность. Случайная и систематическая погрешность. Погрешность косвенных измерений. Правила округления. Смещённая и несмещённая оценка дисперсии. Стандартное отклонение. Квантили и доверительные интервалы (на примере нормального распределения и распределения Стьюдента)

Проверка статистических гипотез. Критерии Стьюдента, Фишера, Пирсона. Проверка распределения на нормальность.

Числа с плавающей запятой и машинный ноль. Оценка погрешности формул численного дифференцирования с выбором оптимального шага.

A2. Теоретические вопросы по методу наименьших квадратов

Метод наименьших квадратов. Линейная регрессия. Вывод формул для коэффициентов регрессии, их доверительных интервалов, стандартной ошибки регрессии. Критерий Стьюдента и Фишера.

Метод наименьших квадратов. Нелинейная регрессия и способы линеаризации (полиномы, функции типа уравнения Аррениуса, уравнений ферментативной кинетики и т.п.). Влияние линеаризации на значение параметров. Преимущества и недостатки линеаризации по сравнению с прямым решением задачи.

Метод наименьших квадратов. Нелинейная регрессия и численные методы. Методы Ньютона, Гаусса-Ньютона и Левенберга-Марквардта, их сравнение друг с другом. Матрица Якоби и оценка доверительных интервалов.

Матрица Якоби в нелинейной регрессии и её роль в оценке погрешностей. Способы её расчёта: аналитическое, численное и автоматическое дифференцирование.

Мультиколлинеарность и обусловленность задачи. Регуляризация по Тихонову и гребневая регрессия (ридж-регрессия). Смещённые и несмещённые оценки параметров модели.

Методы глобальной оптимизации. Метод отжига и генетические алгоритмы. Символьная регрессия. Выбор варианта метода наименьших квадратов для решения конкретной задачи.

Б1. Практическая задача по основам MATLAB

Проверка нормальности распределения

Практическая реализация генетического алгоритма

Практическая реализация метода Монте-Карло для численного интегрирования объёма фигуры

Б2. Практическая задача по физико-химическим расчётам

Реализация метода Ньютона с векторизацией кода на примере расчёта рН раствора (кривые титрования)

Построение p -х и T -х диаграмм «жидкость-пар» двухкомпонентной системы с азеотропом

Построение диаграммы плавкости двухкомпонентной системы с эвтектикой. Оптимизация параметра модели.

Методические материалы для проведения процедур оценивания результатов обучения

Зачет проходит по билетам, каждый из которых включает один теоретический вопрос и одно практическое контрольное задание (ПКЗ). Уровень знаний аспиранта оценивается по каждому вопросу на «отлично», «хорошо», «удовлетворительно», «неудовлетворительно». В случае если на все вопросы был дан ответ, оцененный не ниже, чем «удовлетворительно», аспирант получает общую оценку «зачтено».

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТА ОБУЧЕНИЯ по дисциплине (модулю)				
Оценка Результат	Незачёт (2)	Зачёт (3)	Зачёт (4)	Зачёт (5)
Знания	Отсутствие базовых знаний	Общие, но неглубокие знания, содержащие пробелы	Общие, но не структурированные знания	Сформированные систематические знания
Умения	Отсутствие умений	В целом успешное, но не систематическое умение	В целом успешное, но содержащее отдельные пробелы умение (допускает неточности непринципиального характера)	Успешное и систематическое умение
Навыки (владения)	Отсутствие навыков	Наличие навыков, не всегда верно используемых	В целом, сформированные навыки, но не в активной форме	Сформированные навыки, применяемые при решении поставленных задач