

Стивен Л. Брантон, Дж. Натан Куц

Анализ данных в науке и технике

Data-Driven Science and Engineering

**Machine Learning, Dynamical Systems,
and Control**

Steven L. Brunton, J. Nathan Kutz



CAMBRIDGE
UNIVERSITY PRESS

Анализ данных в науке и технике

**Машинное обучение,
динамические системы и управление**

Стивен Л. Брантон, Дж. Натан Куц



Москва, 2021

УДК 001.5, 004.6
ББК 20, 32.97
Б87

Брантон С. Л., Куц Дж. Н.

Б87 Анализ данных в науке и технике / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2021. – 542 с.: ил.

ISBN 978-5-97060-910-1

Открытия, сделанные на основе анализа данных, совершили революцию в моделировании, прогнозировании поведения и управлении сложными системами. В этой книге приводятся сведения из машинного обучения, инженерной математики и математической физики с целью показать, как моделирование и управление динамическими системами сочетаются с современными методами науки о данных. Рассказывается о многих достижениях в области научных расчетов, которые позволяют применять управляемые данными методы к изучению разнообразных сложных систем, таких как турбулентность, науки о мозге, климатология, эпидемиология, финансы, робототехника и автономные системы.

Книга рассчитана на интересующихся студентов старших курсов и аспирантов первого года обучения инженерных и физических специальностей, в ней рассматривается широкий круг тем и методов на уровне от введения до недавних работ.

УДК 001.5, 004.6
ББК 20, 32.97

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

ISBN 978-1-108-42209-3 (англ.)
ISBN 978-5-97060-910-1 (рус.)

© Steven L. Brunton and J. Nathan Kutz, 2019
© Оформление, издание, перевод,
ДМК Пресс, 2021

Содержание

От издательства	13
Об авторах	14
Предисловие.....	15
Общепотребительные методы оптимизации, уравнения, символы и акронимы	20
Часть I. Понижение размерности и преобразования	31
Глава 1. Сингулярное разложение (SVD)	32
1.1. Общие сведения.....	33
Определение SVD	34
Вычисление SVD	35
Историческая справка	36
Использование в этой книге и предположения о подготовке читателей.....	37
1.2. Аппроксимация матриц	37
Усечение.....	38
Пример: сжатие изображения.....	38
1.3. Математические свойства и манипуляции	41
Интерпретация с привлечением доминирующих корреляций	41
Метод моментальных снимков.....	43
Геометрическая интерпретация	43
Инвариантность SVD относительно унитарных преобразований	45
Левые унитарные преобразования	46
Правые унитарные преобразования	46
1.4. Псевдообращение, метод наименьших квадратов и регрессия.....	47
Одномерная линейная регрессия	49
Полилинейная регрессия	51
Предостережение	53
1.5. Метод главных компонент (PCA)	53
Вычисление	54
Пример: данные с гауссовым шумом.....	55
Пример: данные о раке яичников	57
1.6. Пример: «собственные лица»	58
1.7. Отсечение и выравнивание.....	64
Оптимальный жесткий порог отсечения.....	64
Важность выравнивания данных.....	68

1.8. Рандомизированное сингулярное разложение	71
Рандомизированная линейная алгебра	71
Рандомизированный алгоритм SVD	72
Пример рандомизированного SVD	75
1.9. Тензорные разложения и N -мерные массивы данных	76
Рекомендуемая литература	81

Глава 2. Преобразование Фурье

и вейвлет-преобразование

2.1. Ряд Фурье и преобразование Фурье	83
Скалярные произведения функций и векторов	83
Ряд Фурье	84
Преобразование Фурье	89
2.2. Дискретное преобразование Фурье (ДПФ) и быстрое преобразование Фурье (БПФ)	92
Дискретное преобразование Фурье	93
Быстрое преобразование Фурье	95
Пример БПФ: фильтрация шума	96
Пример БПФ: спектральные производные	98
2.3. Преобразование дифференциальных уравнений в частных производных	100
Уравнение теплопроводности	101
Одностороннее волновое уравнение	103
Уравнение Бюргера	105
2.4. Преобразование Габора и спектрограмма	107
Дискретное преобразование Габора	108
Пример: сигнал с квадратичной частотной модуляцией	108
Пример: «Патетическая соната» Бетховена	110
Принцип неопределенности	112
2.5. Вейвлеты и многомасштабный анализ	113
Дискретное вейвлет-преобразование	115
2.6. Двумерные преобразования и обработка сигналов	116
Двумерное преобразование Фурье для изображений	116
Двумерное вейвлет-преобразование изображений	119
Рекомендуемая литература	122

Глава 3. Разреженность и сжатие измерений

3.1. Разреженность и сжатие	124
Пример: сжатие изображения	125
Почему сигналы допускают сжатие: просторность пространства изображений	127
3.2. Сжатое измерение	128
Заявление об отказе от ответственности	132
Другие формулировки	133
3.3. Примеры сжатых измерений	133
Норма ℓ_1 и разреженные решения недоопределенной системы	134

Восстановление звукового сигнала по разреженным измерениям	135
3.4. Геометрия сжатия	137
Свойство ограниченной изометрии (RIP)	139
Некогерентность и матрицы измерений.....	139
Плохие измерения	140
3.5. Разреженная регрессия.....	140
Отбрасывание выбросов и робастность	141
Отбор признаков и LASSO-регрессия	142
3.6. Разреженное представление.....	146
3.7. Робастный метод главных компонент (RPCA).....	151
3.8. Разреженное размещение датчиков.....	153
Разреженное размещение датчиков для реконструкции	154
Разреженная классификация	158
Рекомендуемая литература	159

Часть II. МАШИННОЕ ОБУЧЕНИЕ И АНАЛИЗ

ДАННЫХ

Глава 4. Регрессия и выбор модели.....

4.1. Классическая аппроксимация кривой	163
Методы наименьших квадратов	163
Линия наименьших квадратов.....	166
Линеаризация данных.....	167
4.2. Нелинейная регрессия и градиентный спуск.....	169
Градиентный спуск	170
Метод переменных направлений	175
4.3. Регрессия и уравнение $Ax = b$: переопределенные и недоопределенные системы.....	176
Переопределенные системы	176
Недоопределенные системы	180
4.4. Оптимизация как краеугольный камень регрессии	183
4.5. Парето-фронт и Lex Parsimoniae	188
Переобучение.....	190
4.6. Выбор модели: перекрестная проверка	191
<i>k</i> -групповая перекрестная проверка.....	195
Перекрестная проверка с контролем по <i>p</i> точкам	197
4.7. Выбор модели: информационный критерий	197
Информационные критерии: AIC и BIC	200
Вычисление AIC и BIC.....	201
Рекомендуемая литература	202

Глава 5. Кластеризация и классификация

5.1. Выделение признаков и добыча данных	204
5.2. Обучение с учителем и без учителя.....	210
5.3. Обучение без учителя: кластеризация методом <i>k</i> средних	214
5.4. Иерархическая кластеризация без учителя: дендрограмма.....	219

5.5. Смесовые модели и EM-алгоритм.....	223
5.6. Обучение с учителем и линейные дискриминанты.....	227
5.7. Метод опорных векторов (SVM)	233
Линейный SVM	233
Нелинейный SVM.....	235
Ядерные методы в сочетании с SVM.....	236
5.8. Решающие деревья и случайные леса	238
Случайные леса.....	243
5.9. 10 лучших алгоритмов по версии Data Mining 2008.....	244
Алгоритм k средних	244
EM-алгоритм (смесовые модели)	245
Метод опорных векторов (SVM).....	245
CART (Classification and Regression Tree – дерево классификации и регрессии)	245
Метод k ближайших соседей (kNN).....	246
Наивная байесовская классификация.....	246
AdaBoost (ансамблевое обучение с усилением)	246
C4.5 (ансамблевое обучение решающих деревьев).....	247
Алгоритм Apriori	247
PageRank	247
Рекомендуемая литература	248

Глава 6. Нейронные сети и глубокое обучение

6.1. Нейронные сети: однослойные сети	250
Однослойная сеть.....	252
6.2. Многослойные сети и функции активации	255
6.3. Алгоритм обратного распространения	260
6.4. Алгоритм стохастического градиентного спуска	264
6.5. Глубокие сверточные нейронные сети.....	267
Сверточные слои	268
Пулинговые слои	269
Полносвязные слои	269
Прореживание	270
6.6. Нейронные сети для динамических систем	272
6.7. Разнообразие нейронных сетей	277
Перцептрон	277
Сети прямого распространения (FF)	277
Рекуррентная нейронная сеть (RNN)	279
Автокодировщик (AE).....	279
Марковская цепь (MC)	280
Сеть Хопфилда (HN).....	280
Машина Больцмана (BM)	280
Ограниченная машина Больцмана (RBM)	281
Сеть глубокого доверия (DBN).....	281
Глубокая сверточная нейронная сеть (DCNN).....	281
Антисверточная сеть (DN).....	281
Глубокая сверточная сеть обратной графики (DCIGN).....	282

Порождающая состязательная сеть (GAN)	282
Машина неустойчивых состояний (LSM).....	282
Машина экстремального обучения (ELM)	283
Сеть с эхо-состояниями (ESN)	283
Глубокая остаточная сеть (DRN).....	283
Сеть Кохонена (KN)	284
Нейронная машина Тьюринга (NTM).....	284
Рекомендуемая литература	284

Часть III. ДИНАМИЧЕСКИЕ СИСТЕМЫ И УПРАВЛЕНИЕ.....

285

Глава 7. Динамические системы, управляемые данными.....

286

7.1. Обзор, мотивация и проблемы	287
Динамические системы	287
Цели и проблемы современной теории динамических систем	291
7.2. Разложение по динамическим модам (DMD)	294
Алгоритм DMD	295
Пример и код	300
Расширения, приложения и ограничения.....	300
7.3. Разреженная идентификация нелинейной динамики (SINDy)	308
Нахождение дифференциальных уравнений в частных производных.....	314
Обобщение SINDy на рациональные нелинейности	316
Применение информационного критерия для выбора модели	319
7.4. Теория оператора Купмана	320
Математическая теория оператора Купмана.....	320
Разложение по модам Купмана и конечные представления.....	324
Примеры погружений Купмана	326
Аналитическое разложение собственных функций в ряд	329
История и недавние достижения.....	331
7.5. Управляемый данными анализ Купмана	332
Расширенный DMD	332
Аппроксимация собственных функций Купмана на основе данных	334
Управляемый данными анализ Купмана и запаздывающие координаты	336
Нейронные сети для погружений Купмана.....	340
Рекомендуемая литература	342

Глава 8. Теория линейного управления

344

Типы управления	345
8.1. Управление с замкнутым контуром обратной связи.....	346
Примеры преимуществ управления с обратной связью.....	348
8.2. Линейные стационарные системы	351
Линеаризация нелинейной динамики	351

Неуправляемая линейная система	352
Управляемая линейная система	354
Системы с дискретным временем	355
Пример: обратный маятник	356
8.3. Управляемость и наблюдаемость	357
Управляемость	357
Наблюдаемость	359
Критерий управляемости РВН	360
Теорема Кэли–Гамильтона и достижимость	361
Грамианы и степень управляемости и наблюдаемости	362
Стабилизируемость и распознаваемость	364
8.4. Оптимальное управление полным состоянием: линейно-квадратичный регулятор (ЛКР)	364
Вывод уравнения Риккати оптимального управления	366
8.5. Оптимальное оценивание полного состояния: фильтр Калмана	369
8.6. Оптимальное управление с использованием датчиков: линейно-квадратичное гауссово управление (ЛКГ)	372
8.7. Практический пример: обратный маятник на тележке	374
Управление маятником на тележке с обратной связью	376
Оценка полного состояния системы маятник–тележка	379
Управление с обратной связью системой маятник–тележка с использованием датчиков	382
8.8. Робастное управление и методы анализа в частотной области	384
Методы в частотной области	384
Качество управления и передаточная функция контура: чувствительность и дополнительная чувствительность	389
Обращение динамики	392
Робастное управление	393
Рекомендуемая литература	396

Глава 9. Сбалансированные модели, пригодные

для управления	397
9.1. Упрощение модели и идентификация системы	397
9.2. Сбалансированное упрощение модели	399
Цель упрощения модели	399
Замена переменных в системах управления	401
Балансирующие преобразования	403
Сбалансирование усечения	407
Вычисление сбалансированных реализаций	408
Пример сбалансированного упрощения модели	413
9.3. Идентификация системы	415
Алгоритм реализации собственной системы	416
Идентификация наблюдателей с помощью фильтра Калмана	419
Комбинация ERA и OKID	423
Рекомендуемая литература	425

Глава 10. Управление на основе данных	426
10.1. Идентификация нелинейной системы для управления.....	427
DMD с управлением.....	428
Нелинейное управление с помощью оператора Купмана.....	430
SINDy с управлением.....	432
Пример управления на основе прогнозирующих моделей (MPC).....	432
10.2. Управление с машинным обучением.....	436
Обучение с подкреплением.....	438
Управление с итеративным обучением.....	439
Генетические алгоритмы.....	439
Генетическое программирование.....	441
Пример: применение генетического алгоритма для настройки ПИД-регулятора.....	443
10.3. Адаптивное управление с поиском экстремума.....	448
Простой пример управления с поиском экстремума.....	452
Пример управления с поиском экстремума в сложной ситуации.....	455
Приложения управления с поиском экстремума.....	456
Рекомендуемая литература.....	458
Часть IV. МОДЕЛИ ПОНИЖЕННОГО ПОРЯДКА	460
Глава 11. Модели пониженного порядка (ROM)	461
11.1. POD для дифференциальных уравнений в частных производных.....	462
Разложение по модам Фурье.....	465
Специальные функции и теория Штурма–Лиувилля.....	466
Понижение размерности.....	467
11.2. Элементы оптимального базиса: собственное ортогональное разложение.....	468
Проекция Галеркина на POD-моды.....	470
Пример: гармонический осциллятор.....	471
11.3. POD и динамика солитонов.....	475
Упрощение солитона ($N = 1$).....	477
Упрощение солитона ($N = 2$).....	479
11.4. POD в непрерывной формулировке.....	480
Квадратурные правила для R: правило трапеций.....	482
Квадратурные правила более высокого порядка.....	483
POD-моды и квадратурные формулы.....	485
11.5. POD с симметриями: повороты и сдвиги.....	486
Сдвиг: распространение волн.....	486
Поворот: спиральные волны.....	488
Рекомендуемая литература.....	492
Глава 12. Интерполяция для ROM	494
12.1. Неполное POD.....	494
Разреженные измерения и реконструкция.....	496

Моды гармонического осциллятора	497
12.2. Ошибка и сходимость неполного POD	501
Случайная выборка и сходимость	501
Неполные измерения и качество реконструкции.....	503
12.3. Неполные измерения: минимизация числа обусловленности	504
Замены числа обусловленности.....	510
12.4. Неполные измерения: максимальная дисперсия.....	512
12.5. POD и дискретный эмпирический метод интерполяции (DEIM) ...	517
POD и DEIM	518
DEIM	519
12.6. Реализация алгоритма DEIM.....	521
Алгоритм QDEIM	523
12.7. Машинное обучение ROM.....	524
Выбор POD-моды	525
Пример: обтекание цилиндра	527
Рекомендуемая литература	529
Глоссарий.....	531
Список литературы.....	538
Предметный указатель	539

От издательства

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв на нашем сайте www.dmkpress.com, зайдя на страницу книги и оставив комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу dmkpress@gmail.com; при этом укажите название книги в теме письма.

Если вы являетесь экспертом в какой-либо области и заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу http://dmkpress.com/authors/publish_book/ или напишите в издательство по адресу dmkpress@gmail.com.

Скачивание исходного кода примеров

Скачать файлы с дополнительной информацией для книг издательства «ДМК Пресс» можно на сайте www.dmkpress.com на странице с описанием соответствующей книги.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы обеспечить высокое качество наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг, мы будем очень благодарны, если вы сообщите о ней главному редактору по адресу dmkpress@gmail.com. Сделав это, вы избавите других читателей от недопонимания и поможете нам улучшить последующие издания этой книги.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательства «ДМК Пресс» и Springer очень серьезно относятся к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконной публикацией какой-либо из наших книг, пожалуйста, пришлите нам ссылку на интернет-ресурс, чтобы мы могли применить санкции.

Ссылку на подозрительные материалы можно прислать по адресу электронной почты dmkpress@gmail.com.

Мы высоко ценим любую помощь по защите наших авторов, благодаря которой мы можем предоставлять вам качественные материалы.

Об авторах

Стивен Л. Брантон – доцент факультета общего машиностроения в Вашингтонском университете. Также является внештатным сотрудником отделения прикладной математики и науки о данных Института eScience. Область его научных интересов охватывает применение науки о данных и машинного обучения к динамическим системам и управлению в области гидрогазодинамики, биомеханики, оптики, энергетических систем и производства. Является автором двух учебников, лауреатом премии армии и ВВС для молодых ученых, получил право преподавания в инженерном колледже Вашингтонского университета и удостоен премии для молодых преподавателей.

Дж. Натан Куц – профессор прикладной математики Вашингтонского университета, был деканом факультета до 2015 года. Также является внештатным профессором отделения электротехники и физики и старшим научным сотрудником отделения науки о данных в Институте eScience. Область научных интересов охватывает сложные системы и анализ данных, конкретно применение методов машинного обучения и динамических систем и управление в разнообразных приложениях. Автор двух учебников, лауреат премии Боинг за отличное преподавание прикладной математики, а также премии CAREER Национального научного фонда США.

Предисловие

Эта книга посвящена растущей области знаний на пересечении методов обработки больших данных, прикладной оптимизации и классических дисциплин инженерной математики и математической физики. Мы готовили данный материал на протяжении ряда лет, в основном для лекций, читаемых студентам старших курсов и аспирантам технических и физических факультетов.

Обычно такие студенты имеют подготовку в области линейной алгебры, дифференциальных уравнений и научных расчетов, а инженеры также знакомы с теорией управления и (или) дифференциальными уравнениями в частных производных. Однако в большинстве учебных программ научно-технических вузов методы обработки данных и (или) оптимизации освещаются слабо или не включены вовсе. С другой стороны, студенты, обучающиеся по специальностям «информатика» и «статистика», плохо знакомы с динамическими системами и теорией управления. Нашей целью было написать введение в прикладную науку о данных для обеих групп. Включенные в книгу методы отбирались по трем критериям: (1) релевантность, (2) простота и (3) общность. Мы стремились представить широкий круг тем от вводного материала до методов, реально применяемых в исследованиях.

Открытие на основе анализа данных революционизировало наши подходы к моделированию, прогнозированию поведения и управлению сложными системами. Самые насущные научно-технические задачи нашего времени не поддаются эмпирическим моделям и выводам, основанным на первопринципах. Все чаще исследователи обращаются к подходам на основе анализа данных при изучении широкого спектра сложных систем, как то: турбулентность, науки о мозге, климатология, эпидемиология, финансы, робототехника, автономные системы. Такие системы обычно являются нелинейными, динамическими, многомасштабными в пространстве и во времени, многомерными и имеют доминирующие паттерны, которые необходимо охарактеризовать и смоделировать, чтобы в конечном итоге обеспечить сбор данных, прогнозирование, оценку и управление. Благодаря современным математическим методам вкупе с невиданной ранее доступностью данных и располагаемыми вычислительными ресурсами мы теперь можем подступить к неприступным до недавнего времени проблемам. Упомянем лишь малую толику новых методов: надежное восстановление изображения по разреженным и зашумленным измерениям случайных пикселей, управление турбулентностью с помощью машинного обучения, оптимальное размещение датчиков и приводов, идентификация допускающих интерпретацию нелинейных динамических систем на основе одних лишь данных и модели пониженного порядка, позволяющие ускорить изучение и оптимизацию систем со сложной многомасштабной физикой.

Движущим началом современной науки о данных является доступность больших и постоянно увеличивающихся объемов данных вследствие заме-

чательных инноваций в области разработки дешевых датчиков, возросших на порядки вычислительных мощностей и практически неограниченной емкости устройств хранения и скорости передачи. Такое изобилие данных открывает перед учеными и инженерами во всех областях новые возможности для изобретений на основе анализа данных; часто в этой связи говорят о четвертой парадигме научного открытия [245]. Эта четвертая парадигма представляет собой естественную кульминацию первых трех: эмпирического эксперимента, аналитического вывода и вычислительного исследования. Интеграция всех трех методик создает новаторскую платформу для новых открытий на основе данных. Этот процесс научного открытия не нов и по сути дела повторяет усилия титанов научной революции: Иоганна Кеплера (1571–1630) и сэра Исаака Ньютона (1642–1727). Оба сыграли ключевую роль в разработке теоретических принципов небесной механики на базе сочетания эмпирических подходов, основанных на анализе данных, и аналитических вычислений. Наука о данных не заменяет математическую физику и технику, но дополняет ее с учетом достижений XXI века, что больше напоминает возрождение, нежели революцию.

Наука о данных сама по себе не нова, ее предложил больше 50 лет назад Джон Тьюки, предвидевший появление науки, в центре внимания которой будет обучение на данных, или *анализ данных* [152]. С тех пор в науке о данных преобладают два разных подхода [78]. Сообщество *машинного обучения* состоит в основном из специалистов по информатике и интересуется в первую очередь разработкой быстрых, масштабируемых и качественных алгоритмов прогнозирования. Сообщество же *статистического обучения*, которое вовсе необязательно во всем противопоставлять первому, больше сосредоточено на факультетах математической статистики и занимается выводом интерпретируемых моделей. Обе методологии могут похвастаться значительными успехами и закладывают математические и вычислительные основания методов науки о данных. Целью ученых и инженеров должно стать использование этих методов для выведения из результатов наблюдений и обсчета моделей (чаще нелинейных), которые правильно улавливают динамику системы и количественно и качественно обобщаются на ненаблюдавшиеся области фазового, параметрического или прикладного пространства. А в этой книге нашей целью будет применение статистических методов и методов машинного обучения к решению технических задач.

РАССМАТРИВАЕМЫЕ ВОПРОСЫ

В книге обсуждается целый ряд ключевых тем. Во-первых, во многих сложных системах присутствуют доминирующие *низкоразмерные паттерны* данных, несмотря на быстрое увеличение разрешающей способности измерений и вычислений. Базовая структура открывает возможность эффективного размещения датчиков и компактного представления для моделирования и управления. Выделение паттернов тесно связано со второй темой: отысканием *преобразований координат*, позволяющих упростить систему.

Действительно, богатая история математической физики вращается вокруг преобразований координат (например, спектральные разложения, преобразование Фурье, обобщенные функции и т. д.), хотя эти методы в большинстве своем были ограничены простой идеализированной геометрией и линейной динамикой. Умение выводить преобразования *на основе данных* открывает возможность обобщить их на новые задачи с более сложной геометрией и граничными условиями. На протяжении всей книги мы будем интересоваться *динамическими системами и управлением*, т. е. применением методов, основанных на анализе данных, к моделированию и управлению систем, изменяющихся во времени. Красной нитью проходит тема *управляемой данными прикладной оптимизации*, поскольку едва ли не каждый рассматриваемый вопрос так или иначе связан с оптимизацией (например, нахождение *оптимальных* низкоразмерных паттернов, *оптимальное* расположение датчиков, *оптимизация* в машинном обучении, *оптимальное* управление и т. д.). И еще одна, даже более фундаментальная тема – большинство данных организовано в массивы для анализа, а широкое развитие численных инструментов линейной алгебры, начиная с 1960-х годов, лежит в основе математических методов матричных разложений и стратегий решения, встречающихся в этой книге.

БЛАГОДАРНОСТИ

Мы в долгу перед многими талантливыми студентами, сотрудниками и коллегами, которые делились ценными замечаниями и предложениями и оказывали нам поддержку. Особенно мы благодарны Джошуа Проктору (Joshua Proctor), который стоял у истоков этой книги и помогал при планировании ее структуры и организации. Мы также извлекли много полезного из бесед с Бингом Брантоном (Bing Brunton), Игорем Мезичем (Igor Mezić), Берндом Ноаком (Bernd Noack) и Сэмом Тайрой (Sam Taira). Эта книга была бы невозможна без помощи наших сотрудников и коллег, исследования которых отражены в тексте.

На протяжении работы над книгой и чтения соответствующих курсов мы получали чрезвычайно ценные отзывы и комментарии от наших замечательных студентов и постдоков: Трэвиса Эшкама (Travis Askham), Майкла Ау-Юнга (Michael Au-Yeung), Цзе Бая (Zhe Bai), Идо Брайта (Ido Bright), Кэтлин Чемпион (Kathleen Champion), Эмили Кларк (Emily Clark), Чарльза Делаханта (Charles Delahunt), Даниэля Дылевски (Daniel Dylewski), Бена Эрричсона (Ben Erichson), Чарли Фислера (Charlie Fiesler), Синь Фу (Xing Fu), Чена Гонга (Chen Gong), Тарена Гормана (Taren Gorman), Джейкоба Гросека (Jacob Grosek), Сета Хирша (Seth Hirsh), Микала Джонсона (Mikala Johnson), Юрики Кайзер (Eurika Kaiser), Мейсона Кема (Mason Kamb), Джеймса Кьюнерта (James Kunert), Бетани Луш (Bethany Lusch), Педро Майа (Pedro Maia), Критики Манохара (Krithika Manohar), Найла Мэнгана (Niall Mangan), Арианы Мендибль (Ariana Mendible), Томаса Морена (Thomas Mohren), Меган Моррисон (Megan Morrison), Маркуса Куэйда (Markus Quade), Сэма Руди (Sam Rudy), Сюзанны Саргсян

(Susanna Sargsyan), Изабель Шерл (Isabel Scherl), Эли Шлизермана (Eli Shlizerman), Джорджа Степанянца (George Stepaniants), Бена Строма (Ben Strom), Чань Суна (Chang Sun), Роя Тэйлора (Roy Taylor), Мегханы Велагар (Meghana Velagar), Джейка Вехолта (Jake Weholt) и Мэтта Уильямса (Matt Williams). Наши студенты подвигли нас на написание этой книги, благодаря им мы каждый день приходим на работу с радостью и волнением.

Мы также благодарны руководителю издательской группы в Cambridge University Press Лоран Коулз (Lauren Cowles), на которую могли положиться на протяжении всего процесса работы.

ОНЛАЙНОВЫЕ МАТЕРИАЛЫ

Мы с самого начала предполагали, что к книге будут прилагаться обширные дополнительные материалы: код, данные, видео, домашние задания и рекомендуемые способы построения курса. Все эти материалы можно найти на сайте databookuw.com.

Код в сети полнее, чем в книге, в частности включен код генерации рисунков, пригодных для публикации. Визуализация данных была поставлена на первое место среди методов науки о данных в опросе «Состояние науки о данных и машинного обучения», проведенном на Kaggle 2017. Поэтому мы настоятельно рекомендуем читателям скачать код с сайта и в полной мере воспользоваться командами построения графиков и диаграмм.

Мы также записали и разместили на YouTube лекции по большинству тем, включенных в книгу. Есть дополнительные видео для студентов, желающих восполнить пробелы в подготовке по научным расчетам и основам прикладной математики. Этот текст задуман одновременно как справочное пособие и источник материалов к нескольким курсам, рассчитанным на студентов разного уровня. Большинство глав самостоятельны, на их основе можно разработать *курсы молодого бойца*, рассчитанные примерно на 10 часов каждый.

КАК ПОЛЬЗОВАТЬСЯ ЭТОЙ КНИГОЙ

Книга рассчитана на начинающих аспирантов и продвинутых студентов старших курсов научных и технических факультетов. Поэтому методы машинного обучения излагаются с азов, но при этом мы предполагаем, что студенты умеют моделировать физические системы с помощью дифференциальных уравнений и решать их с помощью таких программ, как **ode45**. Рассматриваются как начальные вопросы, так и актуальные исследовательские методы. Наша цель – представить цельный взгляд и математический инструментарий для решения научно-технических задач. Но книга может быть также полезна студентам, изучающим информатику и статистику, которые зачастую мало знают о динамических системах и теории управления. На основе представленного материала можно разработать несколько курсов,

программы некоторых из них имеются на сайте книги и включают домашние задания, наборы данных и код.

Прежде всего мы хотели, чтобы книга была интересной, чтобы она вдохновляла, открывала глаза и вооружала знаниями молодых ученых и инженеров. Мы пытались по возможности не слишком усложнять, не жертвуя при этом глубиной и широтой охвата, без которых не может быть никакой исследовательской работы. Многие главы можно было бы развернуть в целые книги, и такие книги есть. Однако мы также стремились к полноте в той мере, в какой этого можно ожидать от книги, посвященной столь обширной и быстро развивающейся области. Мы надеемся, что книга придется вам по вкусу, что вы овладеете всеми описанными в ней методами и измените мир с помощью прикладной науки о данных!

Общеупотребительные методы оптимизации, уравнения, символы и акронимы

НАИБОЛЕЕ РАСПРОСТРАНЕННЫЕ СТРАТЕГИИ ОПТИМИЗАЦИИ

Метод наименьших квадратов (обсуждается в главах 1 и 4) минимизирует сумму квадратов разностей (ошибок) между фактическими данными и предсказаниями модели. В случае линейного метода наименьших квадратов, когда данные аппроксимируются линейной функцией, имеется решение в замкнутой форме, которое можно найти, приравняв к нулю производную ошибки по каждому неизвестному. Этот подход широко используется в технике и прикладных науках для аппроксимации полиномиальными функциями. Применение нелинейного метода наименьших квадратов обычно требует итеративного уточнения путем аппроксимации нелинейного решения линейным на каждой итерации.

Градиентный спуск (обсуждается в главах 4 и 6) – основной метод выпуклой оптимизации в многомерных системах. Для минимизации ошибки вычисляется градиент аппроксимирующей функции. Решение обновляется итеративно путем *спуска с горы* в пространстве решений. Одномерным вариантом градиентного спуска является метод Ньютона–Рафсона. В многомерном пространстве метод часто находит только локальный минимум. Важнейшими инновациями в приложениях больших данных являются стохастический градиентный спуск и алгоритм обратного распространения, благодаря чему оптимизация сводится к самому вычислению градиента.

Чередующийся градиентный спуск (Alternating Descent Method – ADM) (обсуждается в главе 4) позволяет избежать вычисления градиента за счет того, что на каждом шаге производится оптимизация по одной неизвестной. Таким образом, все неизвестные переменные считаются постоянными, за исключением одной, по которой производится линейный поиск (выпуклая оптимизация). Эта переменная обновляется, после чего фиксируется, и то же самое повторяется для другой переменной. На одном шаге итерации пере-

бираются все неизвестные, а сами итерации продолжают до тех пор, пока не будет достигнута желаемая точность.

Расширенный метод Лагранжа (Augmented Lagrange Method – ALM) (обсуждается в главах 3 и 8) – класс алгоритмов для решения задач условной оптимизации. Они похожи на методы штрафования тем, что заменяют задачу оптимизации с ограничениями последовательностью задач без ограничений и прибавляют к целевой функции штрафной член, который играет роль множителя Лагранжа. Расширенный метод Лагранжа – не то же самое, что метод множителей Лагранжа.

Линейное программирование и симплекс-метод – безотказные алгоритмы выпуклой оптимизации. В линейном программировании целевая функция линейно зависит от неизвестных, а ограничениями являются линейные равенства и неравенства. Вычислив область допустимых решений – выпуклый политоп, – алгоритм линейного программирования находит в полиэдре точку, в которой функция принимает наименьшее (или наибольшее) значение, если таковая существует. Симплекс-метод – это конкретная итеративная процедура линейного программирования, которая по заданному опорному допустимому решению пытается найти другое опорное решение, для которого целевая функция принимает меньшее значение, и тем самым производит оптимизацию.

НАИБОЛЕЕ УПОТРЕБИТЕЛЬНЫЕ УРАВНЕНИЯ И СИМВОЛЫ

Линейная алгебра

Линейная система уравнений

$$\mathbf{Ax} = \mathbf{b}. \quad (0.1)$$

Матрица $\mathbf{A} \in \mathbb{R}^{p \times n}$ и вектор $\mathbf{b} \in \mathbb{R}^p$ обычно известны, а вектор $\mathbf{x} \in \mathbb{R}^n$ неизвестен.

Уравнение для собственных значений

$$\mathbf{AT} = \mathbf{T}\mathbf{\Lambda}. \quad (0.2)$$

Столбец ξ_k матрицы \mathbf{T} является собственным вектором матрицы $\mathbf{A} \in \mathbb{C}^{n \times n}$, соответствующим собственному значению λ_k : $\mathbf{A}\xi_k = \lambda_k \xi_k$. Матрица $\mathbf{\Lambda}$ – диагональная матрица, содержащая эти собственные значения, в простейшем случае все n собственных значений различны.

Замена координат

$$\mathbf{x} = \mathbf{\Psi}\mathbf{a}. \quad (0.3)$$

Вектор $\mathbf{x} \in \mathbb{R}^n$ можно записать как $\mathbf{a} \in \mathbb{R}^n$ в системе координат, определенной столбцами матрицы $\mathbf{\Psi} \in \mathbb{R}^{n \times n}$.

Уравнение измерений

$$\mathbf{y} = \mathbf{C}\mathbf{x}. \tag{0.4}$$

Вектор $\mathbf{y} \in \mathbb{R}^p$ является измерением состояния $\mathbf{x} \in \mathbb{R}^n$ в результате применения матрицы измерений $\mathbf{C} \in \mathbb{R}^{p \times n}$.

Сингулярное разложение

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^* \approx \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^*. \tag{0.5}$$

Матрицу $\mathbf{X} \in \mathbb{C}^{n \times m}$ можно разложить в произведение трех матриц $\mathbf{U} \in \mathbb{C}^{n \times n}$, $\mathbf{\Sigma} \in \mathbb{C}^{n \times m}$ и $\mathbf{V} \in \mathbb{C}^{m \times m}$. Матрицы \mathbf{U} и \mathbf{V} унитарные, т. е. $\mathbf{U}\mathbf{U}^* = \mathbf{U}^*\mathbf{U} = \mathbf{I}^{n \times n}$ и $\mathbf{V}\mathbf{V}^* = \mathbf{V}^*\mathbf{V} = \mathbf{I}^{m \times m}$, где $*$ обозначает операцию комплексного сопряжения и транспонирования. Столбцы \mathbf{U} (соответственно \mathbf{V}) ортогональны и называются левыми (соответственно правыми) *сингулярными векторами*. На главной диагонали диагональной матрицы $\mathbf{\Sigma}$ находятся убывающие неотрицательные элементы, называемые *сингулярными значениями*.

Часто \mathbf{X} аппроксимируется матрицей низкого ранга $\tilde{\mathbf{X}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^*$, где $\tilde{\mathbf{U}}$ и $\tilde{\mathbf{V}}$ содержат первые $r \ll n$ столбцов \mathbf{U} и \mathbf{V} соответственно, а $\tilde{\mathbf{\Sigma}}$ – левый верхний блок $\mathbf{\Sigma}$ размера $r \times r$. В контексте пространственных мод, моделей пониженного порядка и размещения датчиков матрица $\tilde{\mathbf{U}}$ часто обозначается буквой $\mathbf{\Psi}$.

Регрессия и оптимизация

Оптимизация переопределенных и недоопределенных линейных систем

$$\operatorname{argmin}_{\mathbf{x}} (\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 + \lambda g(\mathbf{x})) \text{ или} \tag{0.6a}$$

$$\operatorname{argmin}_{\mathbf{x}} g(\mathbf{x}) \text{ при условии } \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \leq \varepsilon. \tag{0.6b}$$

Здесь $g(\mathbf{x})$ – штраф регрессии (со штрафным параметром λ для переопределенных систем). Для переопределенных и недоопределенных систем линейных уравнений $\mathbf{A}\mathbf{x} = \mathbf{b}$, когда решений либо не существует, либо бесконечно много, для нахождения решения нужно задать ограничение или штраф; эта процедура называется *регуляризацией*.

Оптимизация переопределенных и недоопределенных линейных систем

$$\operatorname{argmin}_{\mathbf{x}} (f(\mathbf{A}, \mathbf{x}, \mathbf{b}) + \lambda g(\mathbf{x})) \text{ или} \tag{0.7a}$$

$$\operatorname{argmin}_{\mathbf{x}} g(\mathbf{x}) \text{ при условии } f(\mathbf{A}, \mathbf{x}, \mathbf{b}) \leq \varepsilon. \tag{0.7b}$$

Это обобщение линейной системы на нелинейную систему $f(\cdot)$ с регуляризацией $g(\cdot)$. Такие переопределенные и недоопределенные системы часто решаются методами градиентного спуска.

Композиционная оптимизация для нейронных сетей

$$\operatorname{argmin}_{\mathbf{A}_j} (f_M(\mathbf{A}_M, \dots, f_2(\mathbf{A}_2, (f_1(\mathbf{A}_1, \mathbf{x})) \dots)) + \lambda g(\mathbf{A}_j)). \tag{0.8}$$

Здесь \mathbf{A}_k – матрицы весов связей между k -м и $(k + 1)$ -м слоями нейронной сети. Обычно это сильно недоопределенная система, которая регуляризи-

руется прибавлением $g(\mathbf{A}_j)$. Композиция и регуляризация весьма важны как для порождения выразительных представлений данных, так и для предотвращения переобучения.

Динамические системы и системы пониженного порядка

*Нелинейное обыкновенное дифференциальное уравнение
(динамическая система)*

$$\frac{d}{dt} \mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), t; \boldsymbol{\beta}). \quad (0.9)$$

Вектор $\mathbf{x}(t) \in \mathbb{R}^n$ описывает состояние системы, изменяющейся во времени t , $\boldsymbol{\beta}$ – вектор параметров, а \mathbf{f} – векторное поле. В общем случае \mathbf{f} – липшицева функция, что гарантирует существование и единственность решения.

Система с линейной зависимостью выхода от входа

$$\frac{d}{dt} \mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \quad (0.10a)$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}. \quad (0.10b)$$

Состояние системы представлено вектором $\mathbf{x} \in \mathbb{R}^n$, входы (приводы) – вектором $\mathbf{u} \in \mathbb{R}^q$, а выходы (датчики) – вектором $\mathbf{y} \in \mathbb{R}^p$. Матрицы \mathbf{A} , \mathbf{B} , \mathbf{C} , \mathbf{D} определяют динамику, управляющее воздействие, стратегию работы датчиков и эффект сквозного управления соответственно.

*Нелинейное отображение
(динамические системы с дискретным временем)*

$$\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k). \quad (0.11)$$

Состояние системы на k -й итерации представлено вектором $\mathbf{x}_k \in \mathbb{R}^n$, а \mathbf{F} – потенциально нелинейное отображение. Часто это отображение описывает продвижение итераций во времени, т. е. $\mathbf{x}_k = \mathbf{x}(k\Delta t)$; в таком случае потоковое отображение обозначается $\mathbf{F}_{\Delta t}$.

Операторное уравнение Купмана (с дискретным временем)

$$\mathcal{K}g = g \circ \mathbf{F}_t \Rightarrow \mathcal{K}_t \varphi = \lambda \varphi. \quad (0.12)$$

Линейный оператор Купмана \mathcal{K}_t экстраполирует функции измерения состояния $g(\mathbf{x})$ с помощью потока \mathbf{F}_t . Собственные значения и собственные векторы \mathcal{K}_t обозначаются λ и $\varphi(\mathbf{x})$ соответственно. Оператор \mathcal{K}_t применяется к гильбертову пространству измерений.

Нелинейные дифференциальные уравнения в частных производных (УрЧП)

$$\mathbf{u}_t = \mathbf{N}(\mathbf{u}, \mathbf{u}_x, \mathbf{u}_{xx}, \dots, x, t; \boldsymbol{\beta}). \quad (0.13)$$

Состояние УрЧП описывается вектором \mathbf{u} , \mathbf{N} – нелинейный оператор эволюции, нижние индексы обозначают взятие частных производных, а x и t –

пространственная и временная переменные соответственно. УрЧП параметризуется значениями, собранными в векторе β . Состояние УрЧП \mathbf{u} может быть непрерывной функцией $u(x, t)$, а может быть дискретизировано в нескольких точках пространства, $\mathbf{u}(t) = [u(x_1, t) \ u(x_2, t) \ \dots \ u(x_n, t)]^T \in \mathbb{R}^n$.

Разложение Галеркина

Непрерывное разложение Галеркина имеет вид:

$$u(x, t) \approx \sum_{k=1}^r a_k(t) \psi_k(x). \tag{0.14}$$

Функции $a_k(t)$ – коэффициенты, отражающие временную динамику, а $\psi_k(x)$ – пространственные моды. Для многомерного дискретизированного состояния разложение Галеркина принимает вид $\mathbf{u}(t) \approx \sum_{k=1}^r a_k(t) \Psi_k$. Пространственные моды $\Psi_k \in \mathbb{R}^n$ могут быть столбцами матрицы $\Psi = \tilde{\mathbf{U}}$.

СПИСОК ОБОЗНАЧЕНИЙ

Размерности

- K количество ненулевых элементов K -разреженного вектора \mathbf{s}
- m количество снимков данных (т. е. столбцов \mathbf{X})
- n размерность состояния $\mathbf{x} \in \mathbb{R}^n$
- p размерность измерения, или выходной переменной $\mathbf{y} \in \mathbb{R}^p$
- q размерность выходной переменной $\mathbf{u} \in \mathbb{R}^q$
- r ранг усеченного сингулярного разложения или иной низкоранговой аппроксимации

Скаляры

- s частота в лапласовой области
- t время
- δ скорость обучения в методе градиентного спуска
- Δt временной шаг
- x пространственная переменная
- Δx пространственный шаг
- σ сингулярное значение
- λ собственное значение
- λ параметр разреженности при разреженной оптимизации (раздел 7.3)
- λ множитель Лагранжа (разделы 3.7, 8.4 и 11.4)
- τ порог

Векторы

- \mathbf{a} вектор амплитуд мод \mathbf{x} в базисе Ψ , $\mathbf{a} \in \mathbb{R}^r$
- \mathbf{b} вектор измерений в линейной системе $\mathbf{Ax} = \mathbf{b}$
- \mathbf{b} вектор амплитуд мод в разложении по динамическим модам (раздел 7.2)
- \mathbf{Q} вектор, содержащий функцию потенциала в алгоритме PDE-FIND

- r** вектор невязок
s разреженный вектор $\mathbf{s} \in \mathbb{R}^n$
u регулируемая переменная (главы 8, 9, 10)
u вектор состояния УрЧП (главы 11 и 12)
w экзогенные входы
 \mathbf{w}_d возмущения системы
 \mathbf{w}_n шум измерений
 \mathbf{w}_r опорная траектория
x состояние системы $\mathbf{x} \in \mathbb{R}^n$
 \mathbf{x}_k снимок данных в момент t_k
 \mathbf{x}_j пример данных $j \in Z := \{1, 2, \dots, m\}$ (главы 5 и 6)
 $\tilde{\mathbf{x}}$ упрощенное состояние $\tilde{\mathbf{x}} \in \mathbb{R}^r$, т. е. $\mathbf{x} \approx \tilde{\mathbf{U}}\tilde{\mathbf{x}}$
 $\hat{\mathbf{x}}$ оценка состояния системы
y вектор измерений $\mathbf{y} \in \mathbb{R}^p$
 y_j метка данных $j \in Z := \{1, 2, \dots, m\}$ (главы 5 и 6)
 $\hat{\mathbf{y}}$ оценка измерения выхода
z преобразованное состояние $\mathbf{x} = \mathbf{Tz}$ (главы 8 и 9)
ε вектор ошибок
β бифуркационные параметры
ξ собственный вектор оператора Купмана (разделы 7.4 и 7.5)
ξ разреженный вектор коэффициентов (раздел 7.3)
φ мода в разложении по динамическим модам
ψ мода собственного ортогонального разложения (POD)
Y вектор измерений УрЧП в алгоритме PDE-FIND

Матрицы

- A** матрица системы уравнений, или динамики
 $\hat{\mathbf{A}}$ редуцированная динамика в r -мерном подпространстве POD
 \mathbf{A}_x матричное представление линейной динамики с состоянием \mathbf{x}
 \mathbf{A}_y матричное представление линейной динамики с наблюдаемыми переменными \mathbf{y}
(A, B, C, B) матрицы системы с непрерывным пространством состояний
($\hat{\mathbf{A}}_d, \hat{\mathbf{B}}_d, \hat{\mathbf{C}}_d, \hat{\mathbf{B}}_d$) матрицы системы с дискретным пространством состояний
($\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}, \tilde{\mathbf{B}}$) матрицы пространства состояний системы в новых координатах $\mathbf{z} = \mathbf{T}^{-1}\mathbf{x}$
($\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}, \tilde{\mathbf{B}}$) матрицы пространства состояний упрощенной системы ранга r
B матрица входных данных с приводов
C матрица линейных измерений состояний
C матрица управляемости
F дискретное преобразование Фурье
G матричное представление линейной динамики состояний и входов $[\mathbf{x}^T \mathbf{u}^T]^T$
H матрица Ганкеля
H' матрица Ганкеля с временным сдвигом
I единичная матрица
K матричная форма оператора Купмана (глава 7)
K коэффициент усиления системы управления с замкнутым контуром (глава 8)

- K_f коэффициент усиления фильтра Калмана
 K_r коэффициент усиления линейно-квадратичного регулятора (ЛКР)
 L низкоранговая часть матрицы X (глава 3)
 O матрица наблюдаемости
 P унитарная матрица, применяемая к столбцам X
 Q весовая матрица стоимости отклонений от нулевого состояния в ЛКР (раздел 8.4)
 Q ортогональная матрица в QR-разложении
 R весовая матрица стоимости управляющих воздействий в ЛКР (раздел 8.4)
 R верхнетреугольная матрица в QR-разложении
 S разреженная часть матрицы X (глава 3)
 T матрица собственных векторов (глава 8)
 T замена координат (главы 8 и 9)
 U левые сингулярные векторы X , $U \in \mathbb{R}^{n \times n}$
 \hat{U} левые сингулярные векторы экономичного сингулярного разложения X , $\hat{U} \in \mathbb{R}^{n \times n}$
 \tilde{U} левые сингулярные векторы (POD-моды) усеченного сингулярного разложения X , $\tilde{U} \in \mathbb{R}^{n \times r}$
 V правые сингулярные векторы X , $V \in \mathbb{R}^{m \times m}$
 \tilde{V} правые сингулярные векторы (POD-моды) усеченного сингулярного разложения X , $\tilde{V} \in \mathbb{R}^{m \times r}$
 Σ матрица сингулярных значений X , $\Sigma \in \mathbb{R}^{n \times m}$
 $\hat{\Sigma}$ матрица сингулярных значений экономичного сингулярного разложения X , $\hat{\Sigma} \in \mathbb{R}^{m \times m}$
 $\tilde{\Sigma}$ матрица сингулярных значений усеченного сингулярного разложения X , $\tilde{\Sigma} \in \mathbb{R}^{r \times r}$
 W собственные векторы \tilde{A}
 W_c грамиан управляемости
 W_o грамиан наблюдаемости
 X матрица данных, $X \in \mathbb{R}^{n \times m}$
 X' матрица данных с временным сдвигом, $X' \in \mathbb{R}^{n \times m}$
 Y проекция матрицы X на ортогональный базис в рандомизированном сингулярном разложении (раздел 1.8)
 Y матрица данных наблюдаемых величин, $Y = g(X)$, $Y \in \mathbb{R}^{p \times m}$ (глава 7)
 Y' матрица данных наблюдаемых величин со сдвигом, $Y' = g(X')$, $Y' \in \mathbb{R}^{p \times m}$ (глава 7)
 Z эскиз матрицы для рандомизированного сингулярного разложения, $Z \in \mathbb{R}^{n \times r}$ (раздел 1.8)
 Θ матрица измерений, умноженная на разреживающий базис, $\Theta = C\Psi$ (глава 3)
 Θ матрица функций-кандидатов для SINDy (раздел 7.3)
 Γ матрица производных функций-кандидатов для SINDy (раздел 7.3)
 Ξ матрица коэффициентов функций-кандидатов для SINDy (раздел 7.3)
 Ξ матрица нелинейных снимков для DEIM (раздел 12.5)
 Λ диагональная матрица собственных значений
 Y матрица входных снимков, $Y \in \mathbb{R}^{q \times m}$
 Φ матрица DMD-мод, $\Phi \triangleq X'V'\Sigma^{-1}W$

Ψ ортонормированный базис (например, моды Фурье или POD-моды)

Тензоры

$(\mathcal{A}, \mathcal{B}, \mathcal{M})$ тензоры N -мерных массивов размера $I_1 \times I_2 \times \dots \times I_N$

Нормы

$\|\cdot\|_0$ псевдонорма ℓ_0 вектора \mathbf{x} : количество ненулевых элементов \mathbf{x}

$\|\cdot\|_1$ норма ℓ_1 вектора \mathbf{x} : $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$

$\|\cdot\|_2$ норма ℓ_2 вектора \mathbf{x} : $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n (x_i^2)}$

$\|\cdot\|_2$ норма ℓ_2 матрицы \mathbf{X} : $\|\mathbf{X}\|_2 = \max_x \frac{\|\mathbf{X}\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$

$\|\cdot\|_F$ норма Фробениуса матрицы \mathbf{X} : $\|\mathbf{X}\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m |X_{ij}|^2}$

$\|\cdot\|_*$ ядерная норма матрицы \mathbf{X} : $\|\mathbf{X}\|_* = \text{trace}(\sqrt{\mathbf{X}^* \mathbf{X}}) = \sum_{i=1}^m \sigma_i$ (для $m \leq n$)

$\langle \cdot, \cdot \rangle$ скалярное произведение. Для функций $\langle f(x), g(x) \rangle = \int_{-\infty}^{\infty} f(x)g^*(x)dx$

$\langle \cdot, \cdot \rangle$ скалярное произведение. Для векторов $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^* \mathbf{v}$

Операторы, функции и отображения

\mathcal{F} преобразование Фурье

\mathbf{F} отображение динамической системы с дискретным временем

\mathbf{F}_t дискретное потоковое отображение динамической системы

\mathbf{f} динамическая система с непрерывным временем

G преобразование Габора

\mathbf{G} передаточная функция, отображающая входы на выходы (глава 8)

g скалярная функция измерения \mathbf{x}

\mathbf{g} скалярная функция измерения \mathbf{x}

J функция стоимости для регулирования

ℓ функция потерь в методе опорных векторов (глава 5)

\mathcal{K} оператор Купмана (с непрерывным временем)

\mathcal{K}_t оператор Купмана, ассоциированный с потоковым отображением

\mathcal{L} преобразование Лапласа

\mathbf{L} передаточная функция контура (глава 8)

\mathbf{L} линейное дифференциальное уравнение в частных производных (главы 11 и 12)

\mathbf{N} нелинейное дифференциальное уравнение в частных производных

\mathcal{O} порядок величины

\mathbf{S} функция чувствительности (глава 8)

\mathbf{T} дополнительная функция чувствительности (глава 8)

\mathcal{W} вейвлет-преобразование

μ несогласованность между матрицей измерений \mathbf{C} и базисом Ψ

κ число обусловленности

φ собственная функция Купмана

∇ оператор градиента

$*$ оператор свертки

Наиболее употребительные акронимы

БПФ	быстрое преобразование Фурье
ГО	глубокое обучение
ОДУ	обыкновенное дифференциальное уравнение
СНС	сверточная нейронная сеть
УрЧП	дифференциальное уравнение в частных производных
DMD	разложение по динамическим модам (dynamic mode decomposition)
РСА	метод главных компонент (principal components analysis)
POD	собственное ортогональное разложение (proper orthogonal decomposition)
ROM	модель пониженного порядка (reduced order model)
SVD	сингулярное разложение (singular value decomposition)

Прочие акронимы

ДПФ	дискретное преобразование Фурье
ИНС	искусственная нейронная сеть
ЛДА	линейный дискриминантный анализ
НУШ	нелинейное уравнение Шрёдингера
ОПФ	оконное преобразование Фурье (short time Fourier transform)
ПИД	пропорционально-интегрально-дифференцирующий регулятор
РНС	рекуррентная нейронная сеть
СГС	стохастический градиентный спуск
ADM	метод переменных направлений (alternating directions method)
AIC	информационный критерий Акаике (Akaike information criterion)
ALM	расширенный метод множителей Лагранжа (augmented Lagrange multiplier)
ARMA	авторегрессионное скользящее среднее (autoregressive moving average)
ARMAX	авторегрессионное скользящее среднее с экзогенным входом (autoregressive moving average with exogenous input)
BIC	байесовский информационный критерий (Bayesian information criterion)
BPOD	сбалансированное собственное ортогональное разложение (balanced proper orthogonal decomposition)
ССА	канонический корреляционный анализ (canonical correlation analysis)
CFD	вычислительная гидродинамика (computational fluid dynamics)
CoSaMP	согласованное преследование со сжатой выборкой (compressive sampling matching pursuit)
CWT	непрерывное вейвлет-преобразование (continuous wavelet transform)
DCT	дискретное косинусное преобразование (discrete cosine transform)
DEIM	дискретный эмпирический метод интерполяции (discrete empirical interpolation method)
DMDc	разложение по динамическим модам с управлением (dynamic mode decomposition with control)

DMDc	разложение по динамическим модам с управлением (dynamic mode decomposition with control)
DNS	прямое численное моделирование (direct numerical simulation)
DWT	дискретное вейвлет-преобразование
ECOG	электрокортикография (electrocorticography)
eDMD	расширенное DMD (extended DMD)
EIM	эмпирический метод интерполяции (empirical interpolation method)
EM	математическое ожидание-максимизация (expectation maximization)
EOF	эмпирические ортогональные функции (empirical orthogonal functions)
ERA	алгоритм реализации собственной системы (eigensystem realization algorithm)
ESC	управление с поиском экстремума (extremum-seeking control)
GMM	модель гауссовой смеси (Gaussian mixture model)
HAVOK	ганкелево альтернативное представление оператора Купмана (Hankel alternative view of Koopman)
ICA	метод независимых компонент (independent component analysis)
JL	Джонсона–Линденштраусса (Johnson–Lindenstrauss)
KL	Кульбака–Лейблера (Kullback–Leibler)
KLT	преобразование Карунена–Лоэва (Karhunen–Loève transform)
LAD	наименьшее абсолютное отклонение (least absolute deviations)
LASSO	оператор наименьшего абсолютного сжатия и выборки (least absolute shrinkage and selection operator)
LQE	линейно-квадратичная модель оценки (linear quadratic estimator)
LQG	линейно-квадратичный гауссов регулятор (linear quadratic Gaussian controller)
LQR	линейно-квадратичный регулятор
LTI	линейная стационарная система (linear time invariant system)
MIMO	с несколькими входами и несколькими выходами (multiple input, multiple output)
MLC	управление на основе машинного обучения (machine learning control)
MPE	оценка отсутствующей точки (missing point estimation)
mrDMD	многомасштабное разложение по динамическим модам (multi-resolution dynamic mode decomposition)
NARMAX	нелинейная авторегрессионная модель с экзогенными входами (nonlinear autoregressive model with exogenous inputs)
OKID	идентификация наблюдателей с помощью фильтра Калмана (observer Kalman filter identification)
PBH	критерий Попова–Белевича–Хаутуса (Popov–Belevitch–Hautus test)
PCP	преследование главных компонент (principal component pursuit)
PDE-FIND	функциональная идентификация нелинейной динамики с уравнениями в частных производных (partial differential equation functional identification of nonlinear dynamics)

PDF	функция распределения вероятностей (probability distribution function)
PIV	анемометрия по изображениям частиц (particle image velocimetry)
RIP	свойство ограниченной изометрии (restricted isometry property)
RKHS	гильбертово пространство с воспроизводящим ядром (reproducing kernel Hilbert space)
RPCA	робастный метод главных компонент (robust principal components analysis)
rSVD	рандомизированное SVD (randomized SVD)
SINDy	разреженная идентификация нелинейных систем (sparse identification of nonlinear dynamics)
SISO	с одним входом и одним выходом (single input, single output)
SRC	разреженное представление для классификации (sparse representation for classification)
SSA	анализ сингулярного спектра (singular spectrum analysis)
STLS	последовательный метод наименьших квадратов с порогом (sequential thresholded least-squares)
SVM	метод опорных векторов (support vector machine)
TICA	метод независимых компонент с задержкой (time-lagged independent component analysis)
VAC	вариационный подход к конформационной динамике (variational approach of conformation dynamics)

Часть I



**ПОНИЖЕНИЕ
РАЗМЕРНОСТИ
И ПРЕОБРАЗОВАНИЯ**

Глава 1

Сингулярное разложение (SVD)

Сингулярное разложение (SVD) – одно из самых важных разложений матриц, появившихся в компьютерную эру, оно лежит в основе почти всех методов обработки данных, рассматриваемых в этой книге. SVD – это численно устойчивое разложение матрицы, которое применимо для самых разных целей и при этом гарантированно существует. Мы будем использовать SVD для получения низкоранговых аппроксимаций матриц и для вычисления псевдообращения неквадратных матриц, т. е. нахождения решения системы уравнений вида $\mathbf{Ax} = \mathbf{b}$. Еще одно важное применение SVD находит как алгоритм, лежащий в основе метода главных компонент (principal component analysis – PCA), идея которого – выделение статистически значимых факторов из многомерных данных. Комбинация SVD+PCA применяется к решению широкого круга научно-технических задач.

В каком-то смысле SVD является обобщением быстрого преобразования Фурье (БПФ) – темы следующей главы. Учебники прикладной математики часто начинаются с описания БПФ, поскольку это основа многих классических аналитических и численных результатов. Однако БПФ работает в идеализированной постановке, тогда как SVD – более общая техника, основанная на анализе предъявленных данных. Поскольку эта книга посвящена данным, мы начнем с сингулярного разложения, которое можно рассматривать как базис, созданный *специально* под конкретные данные, в отличие от БПФ, предоставляющего *общий* базис.

Во многих предметных областях сложные системы порождают данные, которые естественно организуются в виде больших матриц или, более общо, массивов. Например, временной ряд, получающийся в результате эксперимента или имитационного моделирования, можно представить в виде матрицы, каждый столбец которой содержит все измерения в один момент времени. Если данные в каждый момент времени многомерные, как при построении трехмерной модели погоды высокой разрешающей способности, то их можно *разгладить*, преобразовав в длинный вектор-столбец, и собрать такие столбцы в большую матрицу. Аналогично значения пикселей полутонового изображения можно сохранить в виде матрицы или преобразовать в длинные векторы-столбцы, которые в совокупности образуют матрицу, представляющую кадры фильма. Примечательно, что данные, порождаемые

такими системами, обычно имеют низкий ранг. Это означает, что можно выделить несколько доминирующих паттернов, которые объясняют многомерные данные. SVD как раз и является численно устойчивым и эффективным методом обнаружения таких паттернов в данных.

1.1. ОБЩИЕ СВЕДЕНИЯ

Сейчас мы вкратце опишем SVD и разовьем интуицию, которая поможет продемонстрировать SVD на ряде мотивирующих примеров. SVD лежит в основе многих других методов, описанных в этой книге, в т. ч. методов классификации в главе 5, разложения по динамическим модам (DMD) в главе 7 и собственного ортогонального разложения (POD) в главе 11. Подробно математические свойства обсуждаются в последующих разделах.

Высокая размерность – самая главная трудность при обработке данных, порождаемых сложными системами. Такие наборы данных могут включать аудио, изображения и видео. Данные могут также генерироваться физической системой, например записи электрической активности мозга или измерения скорости течения жидкости в модели либо в эксперименте. Замечено, что во многих естественно возникающих системах в данных присутствуют преобладающие паттерны, характеризуемые аттрактором или многообразием низкой размерности [252, 251].

Рассмотрим, например, типичное изображение – оно содержит много измерений (пикселей) и, стало быть, является точкой в многомерном векторном пространстве. Однако большинство изображений отлично сжимаются, т. е. существенную информацию можно представить подпространством гораздо более низкой размерности. Сжимаемость изображений будет подробно обсуждаться на страницах этой книги. Сложные гидрогазодинамические системы, например атмосфера Земли или турбулентная спутная струя за кормой автомобиля, также дают убедительные примеры низкоразмерной структуры, скрывающейся в пространстве состояний высокой размерности. Хотя в точных моделях поведения жидкости или газа количество степеней свободы исчисляется миллионами и миллиардами, часто в потоке можно выделить доминирующие когерентные структуры, например периодические завихрения за кормой автомобиля или ураганы в атмосфере.

SVD предлагает систематический способ нахождения низкоразмерной аппроксимации данных высокой размерности в терминах доминирующих паттернов. Эту технику можно назвать *управляемой данными*, поскольку для обнаружения паттернов нужны только данные, без привлечения экспертных знаний или интуиции. Метод SVD обладает численной устойчивостью и дает иерархическое представление данных в новой системе координат, определяемой доминирующими корреляциями внутри данных. Кроме того, гарантируется, что сингулярное разложение, в отличие от спектрального, существует для любой матрицы.

SVD имеет много полезных приложений, помимо понижения размерности данных. Этот метод используется для вычисления псевдообратной матрицы для неквадратных матриц и тем самым позволяет найти решения недоопре-

деленных или переопределенных матричных уравнений вида $\mathbf{Ax} = \mathbf{b}$. Также мы будем использовать SVD для очистки наборов данных от шумов. Не менее важно, что SVD позволяет охарактеризовать входную и выходную геометрию линейного отображения векторных пространств. Все эти приложения будут рассмотрены в данной главе и дадут нам возможность развить интуитивное понимание матриц и данных высокой размерности.

Определение SVD

В общем случае нас интересует анализ большого набора данных $\mathbf{X} \in \mathbb{C}^{n \times m}$:

$$\mathbf{X} = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_m \\ | & | & & | \end{bmatrix}. \quad (1.1)$$

Столбцы $\mathbf{x}_k \in \mathbb{C}^n$ могут представлять измерения, полученные в процессе моделирования или эксперимента. Например, это могут быть изображения, вытянутые в векторы-столбцы длиной, равной количеству пикселей в изображении. Векторы-столбцы могут также представлять состояние нестационарной физической системы, например скорости течения жидкости в выбранных точках, множество измерений электрической активности мозга или состояние метеорологической модели с разрешением 1 км.

Индекс k – это метка, обозначающая k -й набор измерений. Во многих примерах из этой книги \mathbf{X} будет содержать *временные ряды* данных, а $\mathbf{x}_k = \mathbf{x}(k\Delta t)$. Зачастую *размерность состояния* n очень велика – порядка миллионов или миллиардов степеней свободы. Столбцы иногда называют (моментальными) *снимками*, а m обозначает количество снимков в \mathbf{X} . Для многих систем $n \gg m$, поэтому получается *высокая и тонкая* матрица, в отличие от *низкой и толстой* в случае, когда $n \ll m$.

SVD – это однозначно определенное разложение, существующее для любой комплексной матрицы $\mathbf{X} \in \mathbb{C}^{n \times m}$:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*, \quad (1.2)$$

где $\mathbf{U} \in \mathbb{C}^{n \times n}$ и $\mathbf{V} \in \mathbb{C}^{m \times m}$ – *унитарные* матрицы¹ с ортонормированными столбцами, а $\mathbf{\Sigma} \in \mathbb{R}^{n \times m}$ – матрица, диагональные элементы которой вещественны и неотрицательны, а все остальные равны нулю. Здесь символ * обозначает операцию комплексного сопряжения и транспонирования². Ниже мы увидим, что унитарность \mathbf{U} и \mathbf{V} важна и часто используется.

Если $n \geq m$, то на диагонали матрицы $\mathbf{\Sigma}$ расположено не более m ненулевых элементов, и ее можно записать в виде $\mathbf{\Sigma} = \begin{bmatrix} \hat{\mathbf{\Sigma}} \\ \mathbf{0} \end{bmatrix}$. Поэтому \mathbf{X} можно *точно* представить, воспользовавшись *экономной* формой SVD:

¹ Квадратная матрица \mathbf{U} называется унитарной, если $\mathbf{U}\mathbf{U}^* = \mathbf{U}^*\mathbf{U} = \mathbf{I}$.

² Для вещественных матриц эта операция совпадает с обычным транспонированием: $\mathbf{X}^* = \mathbf{X}^T$.

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^* = \left[\hat{\mathbf{U}} \quad \hat{\mathbf{U}}^\perp \right] \begin{bmatrix} \hat{\mathbf{\Sigma}} \\ \mathbf{0} \end{bmatrix} \mathbf{V}^* = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\mathbf{V}^*. \quad (1.3)$$

Полное и экономное SVD показаны на рис. 1.1. На столбцы матрицы $\hat{\mathbf{U}}^\perp$ натянуто векторное пространство, дополнительное и ортогональное к натянутому на столбцы $\hat{\mathbf{U}}$. Столбцы \mathbf{U} называются *левыми сингулярными векторами* \mathbf{X} , а столбцы \mathbf{V} – *правыми сингулярными векторами*. Диагональные элементы $\hat{\mathbf{\Sigma}} \in \mathbb{C}^{m \times m}$ называются *сингулярными значениями*, они расположены в порядке убывания. Ранг \mathbf{X} равен количеству ненулевых сингулярных значений.

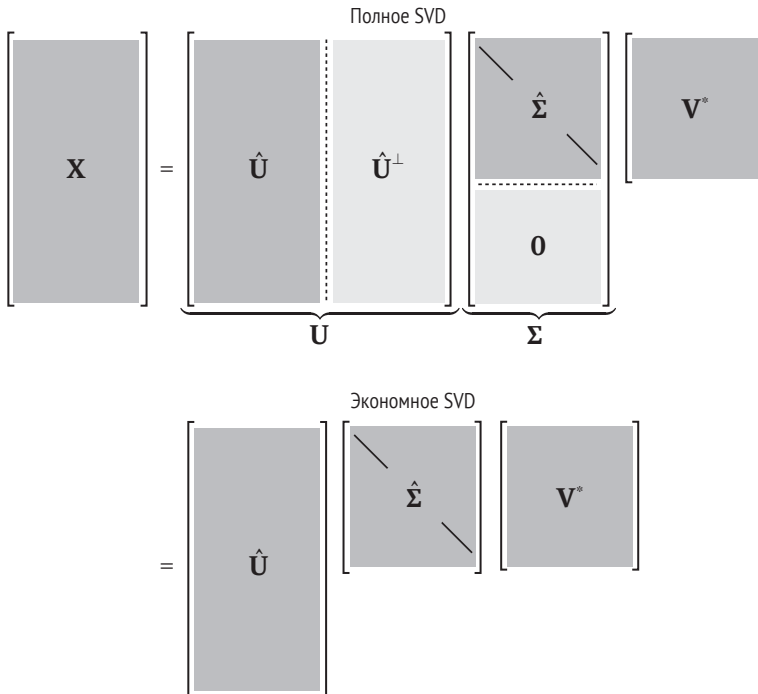


Рис. 1.1 ❖ Схема матриц в полном и экономном SVD

Вычисление SVD

SVD – краеугольный камень вычислительных методов в науке и технике, а численная реализация SVD важна и познавательна с математической точки зрения. Впрочем, большинство стандартных численных реализаций хорошо отработаны, к ним существует простой интерфейс из многих современных языков программирования, что позволяет нам абстрагироваться от деталей вычисления SVD. Как правило, мы будем просто использовать SVD как часть решения более крупной задачи и примем существование эффективных и устойчивых численных алгоритмов как данность. В последующих разделах мы продемон-

стрируем использование SVD на разных языках программирования, а также обсудим большинство стандартных стратегий вычисления и их ограничений. На тему вычисления SVD есть немало важных результатов [212, 106, 211, 292, 238]. Более полное обсуждение вычислительных аспектов можно найти в работе [214]. Для вычисления SVD очень больших матриц все чаще используют рандомизированные алгоритмы, этот вопрос обсуждается в разделе 1.8.

MATLAB. В MATLAB SVD вычисляется прямолинейно:

```
>>X = randn(5,3); % создать случайную матрицу 5×3
>>[U,S,V] = svd(X); % сингулярное разложение
```

Для неквадратных матриц X экономное SVD эффективнее:

```
>>[Uhat,Shat,V] = svd(X,'econ'); % экономное SVD
```

Python

```
>>> import numpy as np
>>> X = np.random.rand(5, 3) % create random data matrix
>>> U, S, V = np.linalg.svd(X,full_matrices=True) % full SVD
>>> Uhat, Shat, Vhat = np.linalg.svd(X, full_matrices=False)
% economy SVD
```

R

```
> X <- replicate(3, rnorm(5))
> s <- svd(X)
> U <- s$u
> S <- diag(s$d)
> V <- s$v
```

Mathematica

```
In:= X=RandomReal[{0,1},{5,3}]
In:= {U,S,V} = SingularValueDecomposition[X]
```

Другие языки

SVD реализовано и на других языках, например Fortran и C++. На самом деле большинство реализаций SVD основаны на библиотеке LAPACK (Linear Algebra Package) [13], написанной на Fortran. В LAPACK подпрограмма вычисления SVD называется **DGESVD**, а она уже обернута функциями C++ в библиотеках **Armadillo** и **Eigen**.

Историческая справка

SVD имеет давнюю и богатую историю, от ранних работ, в которых был заложен теоретический фундамент, до современных исследований по численной устойчивости и эффективности. Отличный исторический обзор приведен в работе Stewart [502], где описан общий контекст и многие важные детали. В этом обзоре много внимания уделено ранним теоретическим работам Бельтрами и Джордана (1873), Сильвестра (1889), Шмидта (1907) и Вейля (1912). Обсуждаются также более поздние работы, включая основополагаю-

щие труды по вычислительной стороне проблемы Голуба с сотрудниками [212, 211]. Кроме того, в современных учебниках [524, 17, 316] имеются прекрасно написанные главы, посвященные SVD.

Использование в этой книге и предположения о подготовке читателей

SVD – основа многих методов понижения размерности. К ним относятся метод главных компонент (PCA) в статистике [418, 256, 257], преобразование Карунена–Лоэва (KLT) [280, 340], эмпирические ортогональные функции (EOF) в изучении климата [344], собственное ортогональное разложение (POD) в гидродинамике [251] и канонический корреляционный анализ (CCA) [131]. Хотя многие из этих методов разрабатывались независимо для разных областей знания, различаются они только способами сбора и предварительной обработки данных. В работе Gerbrands [204] приведено превосходное обсуждение связи между SVD, KLT и PCA.

SVD широко используется для идентификации систем и в теории управления для получения моделей пониженного порядка, сбалансированных в том смысле, что состояния иерархически упорядочены в терминах доступности наблюдению и управляемости с помощью приводов [388].

В этой главе предполагается, что читатель знаком с линейной алгеброй и имеет некоторый опыт применения вычислительных методов. Если необходимо освежить познания, то существует много отличных книг по численной линейной алгебре, в т. ч. с обсуждением SVD, например [524, 17, 316].

1.2. АППРОКСИМАЦИЯ МАТРИЦ

Пожалуй, самым полезным и определяющим свойством SVD является тот факт, что оно дает *оптимальную* низкоранговую аппроксимацию матрицы \mathbf{X} . На самом деле SVD позволяет построить *иерархию* низкоранговых аппроксимаций, поскольку для получения аппроксимации ранга r нужно просто оставить первые r сингулярных значений и векторов, а остальные отбросить.

Шмидт обобщил SVD на пространства функций и доказал теорему, устанавливающую, что усеченное SVD является оптимальной низкоранговой аппроксимацией исходной матрицы \mathbf{X} [476]. Теорема Шмидта была заново открыта в работе Eckart and Young [170], поэтому иногда ее называют теоремой Эккарта–Янга.

Теорема 1 (Eckart–Young [170]). *Оптимальную в смысле наименьших квадратов аппроксимацию \mathbf{X} ранга r дает усеченное SVD $\tilde{\mathbf{X}}$ ранга r :*

$$\operatorname{argmin}_{\tilde{\mathbf{X}} \text{ такое, что } \operatorname{rank}(\tilde{\mathbf{X}})=r} \|\mathbf{X} - \tilde{\mathbf{X}}\|_F = \tilde{\mathbf{U}}\tilde{\Sigma}\tilde{\mathbf{V}}^* \quad (1.4)$$

Здесь $\tilde{\mathbf{U}}$ и $\tilde{\mathbf{V}}$ обозначают матрицы, образованные первыми r столбцами \mathbf{U} и \mathbf{V} , а $\tilde{\Sigma}$ содержит левый верхний блок Σ размера $r \times r$. $\|\cdot\|_F$ – норма Фробениуса.

Согласно этой нотации базис усеченного SVD (и аппроксимирующая матрица $\tilde{\mathbf{X}}$) обозначается $\tilde{\mathbf{X}} = \tilde{\mathbf{U}}\tilde{\Sigma}\tilde{\mathbf{V}}^*$. Поскольку Σ – диагональная матрица, SVD-аппроксимация ранга r может быть представлена суммой r матриц ранга 1:

$$\tilde{\mathbf{X}} = \sum_{k=1}^r \sigma_k \mathbf{u}_k \mathbf{v}_k^* = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^* + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^* + \dots + \sigma_r \mathbf{u}_r \mathbf{v}_r^*. \quad (1.5)$$

Это так называемое *диадическое* суммирование. Для любого ранга r не существует лучшей аппроксимации \mathbf{X} в смысле нормы ℓ_2 , чем усеченная SVD-аппроксимация $\tilde{\mathbf{X}}$. Таким образом, данные высокой размерности хорошо описываются несколькими доминирующими паттернами, определяемыми столбцами $\tilde{\mathbf{U}}$ и $\tilde{\mathbf{V}}$.

Это важное свойство SVD, к которому мы будем много раз возвращаться. Есть многочисленные примеры наборов данных, содержащих измерения высокой размерности, которые приводят к большой матрице \mathbf{X} . Однако в данных часто присутствуют доминирующие паттерны низкой размерности, и базис усеченного SVD $\tilde{\mathbf{U}}$ определяет преобразование координат из пространства измерений высокой размерности в пространство паттернов низкой размерности. В результате уменьшается размер и размерность больших наборов данных, а значит, открывается возможность для визуализации и анализа. Наконец, многие системы, рассматриваемые в этой книге, *динамические* (см. главу 7), а базис SVD дает иерархию мод, характеризующую наблюдаемый аттрактор, на который мы можем спроецировать динамическую систему низкой размерности для получения моделей пониженного порядка (см. главу 12).

Усечение

Усеченное SVD показано на рис. 1.2, где $\tilde{\mathbf{U}}$, $\tilde{\Sigma}$ и $\tilde{\mathbf{V}}$ обозначают усеченные матрицы. Если \mathbf{X} – матрица неполного ранга, то некоторые сингулярные значения в $\tilde{\Sigma}$ могут быть равны нулю, и тогда усеченное SVD остается точным. Однако если r меньше числа ненулевых сингулярных значений (т. е. ранга \mathbf{X}), то усеченное SVD является всего лишь аппроксимацией \mathbf{X} :

$$\mathbf{X} \approx \tilde{\mathbf{U}}\tilde{\Sigma}\tilde{\mathbf{V}}^*. \quad (1.6)$$

Есть много способов выбора ранга усеченной матрицы r , они обсуждаются в разделе 1.7. Если мы требуем, чтобы усеченная матрица содержала все ненулевые сингулярные значения, то равенство $\mathbf{X} = \tilde{\mathbf{U}}\tilde{\Sigma}\tilde{\mathbf{V}}^*$ точное.

Пример: сжатие изображения

Проиллюстрируем идею аппроксимации матриц на простом примере: сжатие изображения. Тема, которая красной нитью проходит через всю книгу, – наличие в больших наборах данных паттернов, благодаря чему возможны низкоранговые представления. Естественные изображения дают простой

и интуитивно понятный пример такой внутренне присущей *сжимаемости*. Полутонное изображение можно рассматривать как вещественную матрицу $\mathbf{X} \in \mathbb{R}^{n \times m}$, где n и m – числа пикселей по вертикали и по горизонтали соответственно¹. В зависимости от базиса представления (пространство пикселей, частотная область в смысле преобразования Фурье, преобразованные с помощью SVD координаты) изображение может иметь очень компактные аппроксимации.

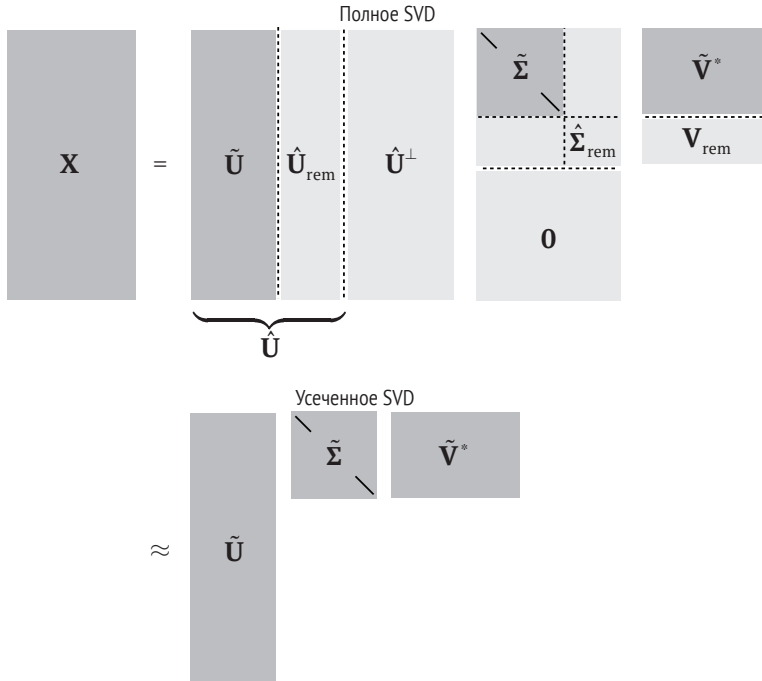


Рис. 1.2 ❖ Схема усеченного SVD.
Нижний индекс «rem» обозначает остаток $\hat{\mathbf{U}}$, $\hat{\Sigma}$ или \mathbf{V} после усечения

Рассмотрим изображение собаки по кличке Мордекай на снегу (рис. 1.3). Его размер 2000×1500 пикселей. Мы можем вычислить SVD этого изображения и нанести на график сингулярные значения (рис. 1.4). На рис. 1.3 показаны приближенные матрицы $\tilde{\mathbf{X}}$ для разных значений r . При $r = 100$ реконструированное изображение вполне точное, а сингулярные значения отражают почти 80 % неоднородности изображения. Усечение SVD приводит к сжатию исходного изображения, поскольку в $\tilde{\mathbf{U}}$, $\tilde{\Sigma}$ и $\tilde{\mathbf{V}}$ нужно хранить только первые 100 столбцов \mathbf{U} и \mathbf{V} плюс первые 100 диагональных элементов Σ .

¹ Размер изображения часто задают, указывая сначала размер по горизонтали, а затем по вертикали, т. е. $\mathbf{X}^T \in \mathbb{R}^{m \times n}$, но мы будем придерживаться противоположного соглашения, совпадающего с общепринятым порядком обозначения размера матрицы.

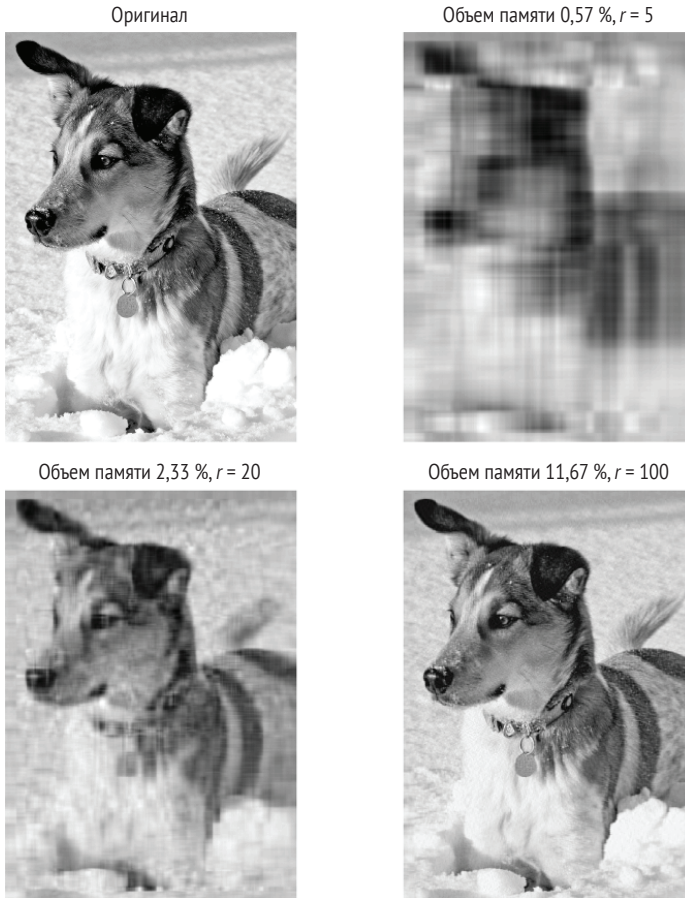


Рис. 1.3 ❖ Сжатие изображения собаки Мордекая на снегу путем усечения SVD с различными значениями ранга r . Разрешение исходного изображения 2000×1500

Сначала загрузим изображение:

```
A=imread('..\\DATA\\dog.jpg');
X=double(rgb2gray(A)); % преобразовать RGB в полутоновое, 256 бит->double.
nx = size(X,1); ny = size(X,2);
imagesc(X), axis off, colormap gray
```

и вычислим SVD:

```
[U,S,V] = svd(X);
```

Затем вычислим приближенную матрицу, используя усеченные SVD с различными рангами ($r = 5, 20, 100$):

```
for r=[5 20 100]; % Truncation value
    Xapprox = U(:,1:r)*S(1:r,1:r)*V(:,1:r)'; % Approx. image
```



```
figure, imagesc(Xapprox), axis off
title(['r=', num2str(r, '%d'), '']);
end
```

Наконец, построим графики сингулярных значений и суммарной энергии, изображенные на рис. 1.4.

```
subplot(1,2,1), semilogy(diag(S), 'k')
subplot(1,2,2), plot(cumsum(diag(S))/sum(diag(S)), 'k')
```

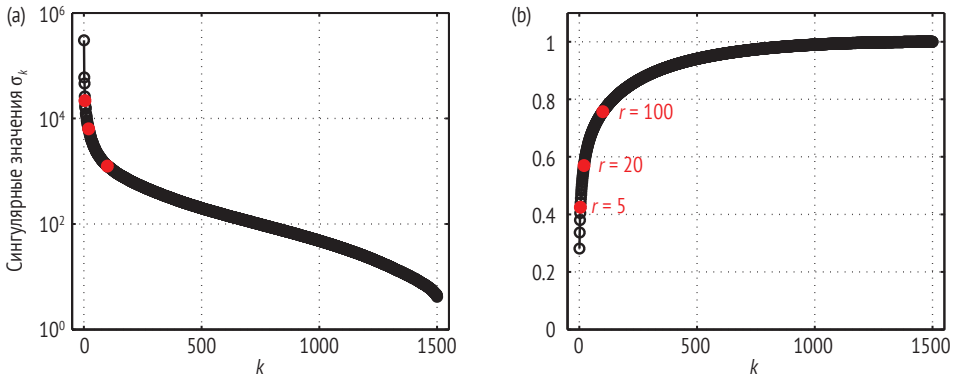


Рис. 1.4 ❖ (a) Сингулярные значения σ_k .
(b) Суммарная энергия в первых k модах

1.3. МАТЕМАТИЧЕСКИЕ СВОЙСТВА И МАНИПУЛЯЦИИ

Опишем важные математические свойства SVD, в т. ч. геометрические интерпретации унитарных матриц \mathbf{U} и \mathbf{V} , а также обсудим SVD в терминах доминирующих корреляций, присутствующих в данных \mathbf{X} . Связь между SVD и корреляциями данных мы объясним в разделе 1.5, посвященном методу главных компонент.

Интерпретация с привлечением доминирующих корреляций

Сингулярное разложение тесно связано с задачей о собственных значениях, в которой фигурируют корреляционные матрицы $\mathbf{X}\mathbf{X}^*$ и $\mathbf{X}^*\mathbf{X}$, показанные на рис. 1.5 для конкретного изображения, а на рис. 1.6 и 1.7 для матриц общего вида. Подставив (1.3) в $\mathbf{X}\mathbf{X}^*$ и $\mathbf{X}^*\mathbf{X}$, получаем:

$$\mathbf{X}\mathbf{X}^* = \mathbf{U} \begin{bmatrix} \hat{\Sigma} & \\ & \mathbf{0} \end{bmatrix} \mathbf{V}^* \mathbf{V} \begin{bmatrix} \hat{\Sigma} & \mathbf{0} \end{bmatrix} \mathbf{U}^* = \mathbf{U} \begin{bmatrix} \hat{\Sigma}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{U}^*; \quad (1.7a)$$

$$\mathbf{X}^* \mathbf{X} = \mathbf{V} \begin{bmatrix} \hat{\Sigma} & \mathbf{0} \end{bmatrix} \mathbf{U}^* \mathbf{U} \begin{bmatrix} \hat{\Sigma} \\ \mathbf{0} \end{bmatrix} \mathbf{V}^* = \mathbf{V} \hat{\Sigma}^2 \mathbf{V}^*. \quad (1.7b)$$

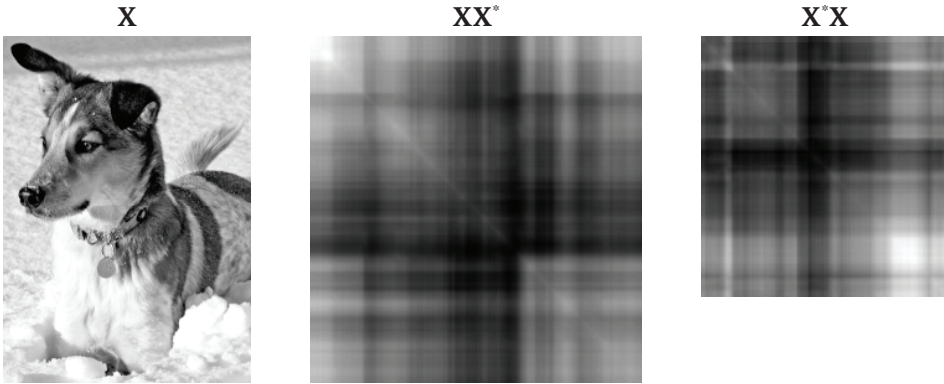


Рис. 1.5 ❖ Корреляционные матрицы $\mathbf{X}\mathbf{X}^*$ и $\mathbf{X}^*\mathbf{X}$ для матрицы \mathbf{X} , полученной из изображения собаки. Заметим, что обе корреляционные матрицы симметричны

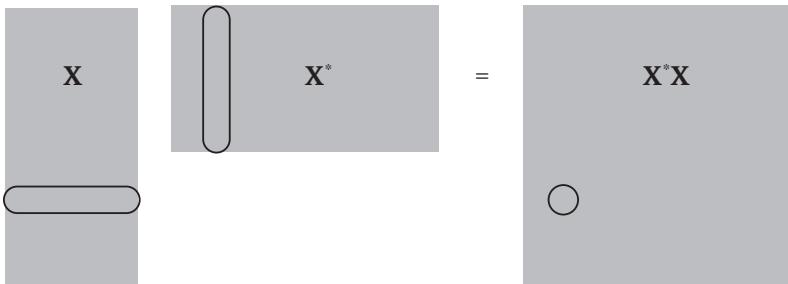


Рис. 1.6 ❖ Корреляционная матрица $\mathbf{X}\mathbf{X}^*$ получена взятием скалярных произведений строк \mathbf{X}

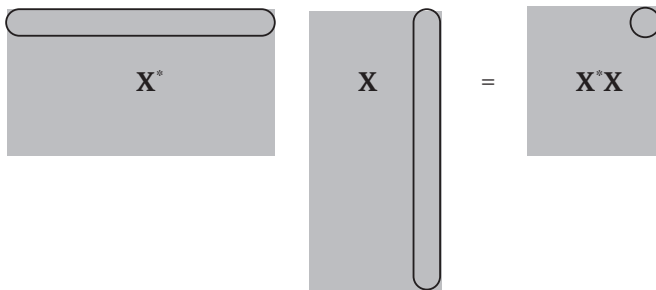


Рис. 1.7 ❖ Корреляционная матрица $\mathbf{X}^*\mathbf{X}$ получена взятием скалярных произведений столбцов \mathbf{X}

Учитывая, что \mathbf{U} и \mathbf{V} унитарны, \mathbf{U} , $\hat{\Sigma}$ и \mathbf{V} являются решениями следующих задач на нахождение собственных значений:

$$\mathbf{X}\mathbf{X}^*\mathbf{U} = \mathbf{U} \begin{bmatrix} \hat{\Sigma}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}; \quad (1.8a)$$

$$\mathbf{X}^*\mathbf{X}\mathbf{V} = \mathbf{V}\hat{\Sigma}^2. \quad (1.8b)$$

Иными словами, любое ненулевое сингулярное значение \mathbf{X} является положительным квадратным корнем из какого-то собственного значения матриц $\mathbf{X}^*\mathbf{X}$ и $\mathbf{X}\mathbf{X}^*$, которые имеют одинаковые множества собственных значений. Отсюда следует, что если \mathbf{X} самосопряженная (т. е. $\mathbf{X} = \mathbf{X}^*$), то сингулярные значения \mathbf{X} равны абсолютным величинам собственных значений \mathbf{X} .

В результате мы получаем интуитивно понятную интерпретацию SVD: столбцы \mathbf{U} являются собственными векторами корреляционной матрицы $\mathbf{X}\mathbf{X}^*$, а столбцы \mathbf{V} – собственными векторами $\mathbf{X}^*\mathbf{X}$. Мы упорядочиваем сингулярные значения в порядке убывания абсолютной величины, поэтому порядок столбцов \mathbf{U} отражает, какую часть корреляции между столбцами \mathbf{X} они улавливают; аналогично \mathbf{V} улавливает корреляцию между строками \mathbf{X} .

Метод моментальных снимков

На практике часто невозможно построить матрицу $\mathbf{X}\mathbf{X}^*$, поскольку размерность состояния n слишком велика; что уж говорить о нахождении собственных значений. Если \mathbf{X} состоит из миллиона элементов, то число элементов $\mathbf{X}\mathbf{X}^*$ равно триллиону. В 1987 году Сирович (Sirovich) заметил, что можно не вычислять эту большую матрицу, а найти первые m столбцов \mathbf{U} с помощью метода, получившего название «метод (моментальных) снимков» [490].

Вместо того чтобы вычислять спектральное разложение $\mathbf{X}\mathbf{X}^*$ для получения левых сингулярных векторов \mathbf{U} , мы вычисляем только спектральное разложение гораздо меньшей и простой для работы матрицы $\mathbf{X}^*\mathbf{X}$. Затем из (1.8b) находим \mathbf{V} и $\hat{\Sigma}$. Если у $\hat{\Sigma}$ имеются нулевые сингулярные значения, то мы оставляем только ее часть $\tilde{\Sigma}$, соответствующую r ненулевым значениям, и соответствующие столбцы $\tilde{\mathbf{V}}$ матрицы \mathbf{V} . Зная эти матрицы, можно следующим образом аппроксимировать матрицу $\tilde{\mathbf{U}}$, состоящую из первых r столбцов \mathbf{U} :

$$\tilde{\mathbf{U}} = \mathbf{X}\tilde{\mathbf{V}}\tilde{\Sigma}^{-1}. \quad (1.9)$$

Геометрическая интерпретация

Столбцы матрицы \mathbf{U} образуют ортонормированный базис пространства столбцов \mathbf{X} . Аналогично столбцы \mathbf{V} образуют ортонормированный базис пространства строк \mathbf{X} . Если столбцы \mathbf{X} содержат пространственные измерения в разные моменты времени, то \mathbf{U} кодирует пространственные паттерны, а \mathbf{V} – временные паттерны.

Особенно полезным делает SVD тот факт, что \mathbf{U} и \mathbf{V} – унитарные матрицы, так что $\mathbf{U}\mathbf{U}^* = \mathbf{U}^*\mathbf{U} = \mathbf{I}_{n \times n}$ и $\mathbf{V}\mathbf{V}^* = \mathbf{V}^*\mathbf{V} = \mathbf{I}_{m \times m}$. Это означает, что для решения системы уравнений с матрицей \mathbf{U} или \mathbf{V} нужно просто умножить обе части на транспонированную матрицу. Сложность этой операции составляет $O(n^2)$, в отличие от традиционных методов обращения матрицы общего вида, имеющих сложность $O(n^3)$. Как отмечено в предыдущем разделе и в работе [57], SVD тесно связано со спектральными свойствами компактных самосопряженных операторов $\mathbf{X}\mathbf{X}^*$ и $\mathbf{X}^*\mathbf{X}$.

Сингулярное разложение \mathbf{X} можно геометрически интерпретировать, рассмотрев отображение гиперсферы $S^{n-1} \triangleq \{\mathbf{x} \mid \|\mathbf{x}\|_2 = 1\} \subset \mathbb{R}^n$ в эллипсоид, $\{\mathbf{y} \mid \mathbf{y} = \mathbf{X}\mathbf{x} \text{ для } \mathbf{x} \in S^{n-1}\} \subset \mathbb{R}^m$ посредством умножения на \mathbf{X} . Графически это показано на рис. 1.8 для сферы в \mathbb{R}^3 и отображения посредством умножения на \mathbf{X} с тремя ненулевыми сингулярными значениями. Поскольку умножение на матрицу – линейное отображение, достаточно знать, как оно ведет себя на единичной сфере, чтобы вычислить образ любого вектора.

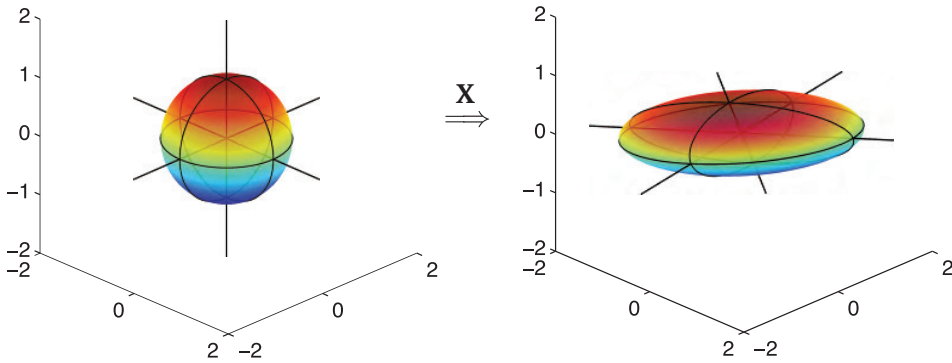


Рис. 1.8 ❖ Геометрическая иллюстрация SVD как отображения сферы в \mathbb{R}^n в эллипсоид в \mathbb{R}^m

Для частного случая на рис. 1.8 мы строим матрицу \mathbf{X} из трех матриц поворота \mathbf{R}_x , \mathbf{R}_y и \mathbf{R}_z и четвертой матрицы, описывающей растяжение вдоль главных осей:

$$\mathbf{X} = \underbrace{\begin{bmatrix} \cos(\theta_3) & -\sin(\theta_3) & 0 \\ \sin(\theta_3) & \cos(\theta_3) & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{R}_z} \underbrace{\begin{bmatrix} \cos(\theta_2) & 0 & \sin(\theta_2) \\ 0 & 1 & 0 \\ -\sin(\theta_2) & 0 & \cos(\theta_2) \end{bmatrix}}_{\mathbf{R}_y} \times \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_1) & -\sin(\theta_1) \\ 0 & \sin(\theta_1) & \cos(\theta_1) \end{bmatrix}}_{\mathbf{R}_y} \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}.$$

В этом случае $\theta_1 = \pi/15$, $\theta_2 = -\pi/9$ и $\theta_3 = -\pi/20$, а $\sigma_1 = 3$, $\sigma_2 = 1$, $\sigma_3 = 0.5$. Матрицы поворота не коммутируют, поэтому порядок поворотов имеет значение. Если одно из сингулярных значений равно нулю, то соответствующая размерность исчезает и эллипсоид становится фигурой в подпространстве меньшей размерности. Произведение $\mathbf{R}_x \mathbf{R}_y \mathbf{R}_z$ является унитарной матрицей \mathbf{U} в сингулярном разложении \mathbf{X} . Матрица \mathbf{V} в данном случае единичная.

Листинг 1.1 ❖ Построение матриц поворота

```
theta = [pi/15; -pi/9; -pi/20];
Sigma = diag([3; 1; 0.5]);           % масштабировать x, y, z
Rx = [1 0 0;                       % поворот вокруг оси x
      0 cos(theta(1)) -sin(theta(1));
      0 sin(theta(1)) cos(theta(1))];
Ry = [cos(theta(2)) 0 sin(theta(2)); % поворот вокруг оси y
      0 1 0;
      -sin(theta(2)) 0 cos(theta(2))];
Rz = [cos(theta(3)) -sin(theta(3)) 0; % поворот вокруг оси z
      sin(theta(3)) cos(theta(3)) 0;
      0 0 1];
X = Rz*Ry*Rx*Sigma;                % повернуть и масштабировать
```

Листинг 1.2 ❖ Изображение сферы

```
[x,y,z] = sphere(25);
h1=surf(x,y,z);
```

Листинг 1.3 ❖ Отображение сферы путем умножения на \mathbf{X} и изображение получившегося эллипсоида

```
xR = 0*x; yR = 0*y; zR = 0*z;
for i=1:size(x,1)
    for j=1:size(x,2)
        vecR = X*[x(i,j); y(i,j); z(i,j)];
        xR(i,j) = vecR(1);
        yR(i,j) = vecR(2);
        zR(i,j) = vecR(3);
    end
end
h2=surf(xR,yR,zR,z);               % координата z сферы задает цвет
```

Инвариантность SVD относительно унитарных преобразований

У SVD есть полезное свойство: если умножить матрицу данных \mathbf{X} слева или справа на унитарную матрицу, то члены сингулярного разложения не изменятся, за исключением левой или правой унитарной матрицы \mathbf{U} или \mathbf{V} соответственно. У этого факта есть важные следствия, поскольку дискретное преобразование Фурье (ДПФ, см. главу 2) \mathcal{F} является унитарным преобразо-

ванием, а это значит, что SVD матрицы $\hat{\mathbf{X}} = \mathcal{F}\mathbf{X}$ совпадает с SVD матрицы \mathbf{X} с тем отличием, что моды $\hat{\mathbf{U}}$ будут дискретными преобразованиями Фурье мод \mathbf{U} : $\hat{\mathbf{U}} = \mathcal{F}\mathbf{U}$. Кроме того, инвариантность SVD относительно унитарных преобразований позволяет использовать сжатые измерения для реконструкции мод SVD, разреженных в некотором базисе преобразования (см. главу 3).

Инвариантность SVD относительно унитарных преобразований геометрически очевидна, поскольку унитарное преобразование лишь поворачивает векторы в пространстве, но не изменяет их скалярные произведения и структуру корреляций. Будем обозначать левое унитарное преобразование \mathbf{C} , так что $\mathbf{Y} = \mathbf{C}\mathbf{X}$, а правое унитарное преобразование \mathbf{P}^* , так что $\mathbf{Y} = \mathbf{X}\mathbf{P}^*$. SVD матрицы \mathbf{X} будем обозначать $\mathbf{U}_X \Sigma_X \mathbf{V}_X^*$, а SVD матрицы \mathbf{Y} будет равно $\mathbf{U}_Y \Sigma_Y \mathbf{V}_Y^*$.

Левые унитарные преобразования

Сначала рассмотрим левое унитарное преобразование \mathbf{X} : $\mathbf{Y} = \mathbf{C}\mathbf{X}$. Вычисление корреляционной матрицы $\mathbf{Y}^* \mathbf{Y}$ дает

$$\mathbf{Y}^* \mathbf{Y} = \mathbf{X}^* \mathbf{C}^* \mathbf{C} \mathbf{X} = \mathbf{X}^* \mathbf{X}. \quad (1.10)$$

Спроецированные данные имеют такой же спектральный состав, т. е. \mathbf{V}_X и Σ_X не изменяются. Применив метод снимков для реконструкции \mathbf{U}_Y , находим

$$\mathbf{U}_Y = \mathbf{Y} \mathbf{V}_X \Sigma_X^{-1} = \mathbf{C} \mathbf{X} \mathbf{V}_X \Sigma_X^{-1} = \mathbf{C} \mathbf{U}_X. \quad (1.11)$$

Таким образом, $\mathbf{U}_Y = \mathbf{C} \mathbf{U}_X$, $\Sigma_Y = \Sigma_X$ и $\mathbf{V}_Y = \mathbf{V}_X$. Тогда SVD матрицы \mathbf{Y} равно:

$$\mathbf{Y} = \mathbf{C} \mathbf{X} = \mathbf{C} \mathbf{U}_X \Sigma_X \mathbf{V}_X^*. \quad (1.12)$$

Правые унитарные преобразования

Для правого унитарного преобразования $\mathbf{Y} = \mathbf{X}\mathbf{P}^*$ корреляционная матрица $\mathbf{Y}^* \mathbf{Y}$ равна

$$\mathbf{Y}^* \mathbf{Y} = \mathbf{P} \mathbf{X}^* \mathbf{X} \mathbf{P}^* = \mathbf{P} \mathbf{V}_X \Sigma_X^2 \mathbf{V}_X^* \mathbf{P}^* \quad (1.13)$$

и имеет такое спектральное разложение:

$$\mathbf{Y}^* \mathbf{Y} \mathbf{P} \mathbf{V}_X = \mathbf{P} \mathbf{V}_X \Sigma_X^2. \quad (1.14)$$

Таким образом, $\mathbf{V}_Y = \mathbf{P} \mathbf{V}_X$ и $\Sigma_Y = \Sigma_X$. Мы можем воспользоваться методом снимков для реконструкции \mathbf{U}_Y :

$$\mathbf{U}_Y = \mathbf{Y} \mathbf{P} \mathbf{V}_X \Sigma_X^{-1} = \mathbf{X} \mathbf{V}_X \Sigma_X^{-1} = \mathbf{U}_X. \quad (1.15)$$

Следовательно, $\mathbf{U}_Y = \mathbf{U}_X$, и SVD матрицы \mathbf{Y} можно записать в виде:

$$\mathbf{Y} = \mathbf{X} \mathbf{P}^* = \mathbf{U}_X \Sigma_X \mathbf{V}_X^* \mathbf{P}^*. \quad (1.16)$$

1.4. ПСЕВДООБРАЩЕНИЕ, МЕТОД НАИМЕНЬШИХ КВАДРАТОВ И РЕГРЕССИЯ

Многие физические системы можно представить линейной системой уравнений

$$Ax = b, \tag{1.17}$$

где матрица ограничений A и вектор b известны, а вектор x неизвестен. Если A – квадратная обратимая матрица (т. е. определитель A не равен нулю), то существует единственное решение x для любого b . Но если A сингулярная или прямоугольная, то может существовать одно, ни одного или бесконечно много решений в зависимости от конкретного b и пространств столбцов и строк A .

Сначала рассмотрим *недоопределенную систему*, т. е. случай, когда $A \in \mathbb{C}^{n \times m}$ и $n \ll m$ (A – низкая и толстая матрица) – уравнений меньше, чем неизвестных. Такая система, скорее всего, будет иметь полный столбцовый ранг, поскольку число столбцов много больше, чем требуется для линейно независимого базиса¹. В общем случае, если низкая и толстая матрица A имеет полный столбцовый ранг, то для каждого b существует бесконечно много решений x . Такая система называется *недоопределенной*, потому что элементов b недостаточно, чтобы определить вектор x высокой размерности.

Точно так же рассмотрим *переопределенную систему*, когда $n \gg m$ (высокая и тощая матрица), т. е. уравнений больше, чем неизвестных. Эта матрица не может иметь полного столбцового ранга, поэтому гарантируется, что существуют такие векторы b , для которых нет ни одного решения x . На самом деле решение x существует, только если b принадлежит пространству столбцов A , т. е. $b \in \text{col}(A)$.

Технически могут существовать b , при которых имеется бесконечно много решений x для высокой и тощей матрицы A , равно как и такие b , при которых не существует ни одного решения для низкой и толстой матрицы. Пространство решений системы (1.17) определяется четырьмя фундаментальными подпространствами $A = \tilde{U}\tilde{\Sigma}\tilde{V}^*$, где ранг r выбран так, чтобы все ненулевые сингулярные значения были включены:

- пространство столбцов, $\text{col}(A)$, натянутое на столбцы A ; оно называется также *областью значений*. Пространство столбцов A совпадает с пространством столбцов \tilde{U} ;
- ортогональное дополнение к $\text{col}(A)$ обозначается $\text{ker}(A^*)$ и совпадает с пространством столбцов матрицы \tilde{U}^\perp на рис. 1.1;
- пространство строк, $\text{row}(A)$, натянутое на строки A и совпадающее с линейной оболочкой столбцов \tilde{V} . Имеет место равенство $\text{row}(A) = \text{col}(A^*)$;

¹ Легко построить вырожденные примеры низкой и толстой матрицы неполного столбцового ранга, например $A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$.

- ядерное пространство, $\ker(\mathbf{A})$, являющееся ортогональным дополнением к $\text{row}(\mathbf{A})$ и называемое также *нуль-пространством*, или *ядром*. Ядро – это подпространство, состоящее из векторов, которые \mathbf{A} отображает в нуль, т. е. $\mathbf{Ax} = \mathbf{0}$; оно совпадает с $\text{col}(\hat{\mathbf{V}}^\perp)$.

Точнее, если $\mathbf{b} \in \text{col}(\mathbf{A})$ и $\dim(\ker(\mathbf{A})) \neq 0$, то существует бесконечно много решений \mathbf{x} . Заметим, что условие $\dim(\ker(\mathbf{A})) \neq 0$ гарантированно выполняется для низкой и толстой матрицы. Аналогично, если $\mathbf{b} \notin \text{col}(\mathbf{A})$, то решений не существует, и система уравнений (1.17) называется *несовместной*.

Описанные выше фундаментальные подпространства обладают следующими свойствами:

$$\text{col}(\mathbf{A}) \oplus \ker(\mathbf{A}^*) = \mathbb{R}^n \quad (1.18a)$$

$$\text{col}(\mathbf{A}^*) \oplus \ker(\mathbf{A}) = \mathbb{R}^n. \quad (1.18b)$$

Замечание 1. *Имеется обширная литература по теории случайных матриц, в которой перечисленные выше утверждения почти всегда верны, т. е. верны с высокой вероятностью. Например, крайне маловероятно, что система $\mathbf{Ax} = \mathbf{b}$ имеет решение для случайно выбранных матрицы $\mathbf{A} \in \mathbb{R}^{n \times m}$ и вектора $\mathbf{b} \in \mathbb{R}^n$, где $n \gg m$, т. к. мало шансов, что \mathbf{b} принадлежит пространству столбцов \mathbf{A} . Эти свойства случайных матриц будут играть важную роль в теории сжатых измерений (см. главу 3).*

В переопределенном случае, когда решений не существует, нам часто хотелось бы найти вектор \mathbf{x} , который минимизирует сумму квадратов ошибок $\|\mathbf{Ax} - \mathbf{b}\|_2^2$, он называется решением с *наименьшей среднеквадратической ошибкой*. Заметим, что такое решение минимизирует также величину $\|\mathbf{Ax} - \mathbf{b}\|_2$. В недоопределенном случае, когда существует бесконечно много решений, часто требуется найти решение \mathbf{x} , для которого $\mathbf{Ax} = \mathbf{b}$, а норма $\|\mathbf{x}\|_2$ минимальна; оно называется решением с *минимальной нормой*.

SVD – общепризнанный метод решения этих важных задач оптимизации. Прежде всего если подставить точное усеченное SVD $\mathbf{A} = \tilde{\mathbf{U}}\tilde{\Sigma}\tilde{\mathbf{V}}^*$ вместо \mathbf{A} , то каждую из матриц $\tilde{\mathbf{U}}$, $\tilde{\Sigma}$ и $\tilde{\mathbf{V}}^*$ можно будет поочередно «обратить», что даст левую псевдообратную матрицу Мура–Пенроуза [425, 426, 453, 572] \mathbf{A}^\dagger :

$$\mathbf{A}^\dagger \triangleq \tilde{\mathbf{V}}\tilde{\Sigma}^{-1}\tilde{\mathbf{U}}^* \Rightarrow \mathbf{A}^\dagger\mathbf{A} = \mathbf{I}^{m \times m}. \quad (1.19)$$

Ее можно использовать для нахождения решений системы (1.17) с наименьшей среднеквадратической ошибкой и с минимальной нормой:

$$\mathbf{A}^\dagger\mathbf{A}\tilde{\mathbf{x}} = \mathbf{A}^\dagger\mathbf{b} \Rightarrow \tilde{\mathbf{x}} = \tilde{\mathbf{V}}\tilde{\Sigma}^{-1}\tilde{\mathbf{U}}^*\mathbf{b}. \quad (1.20)$$

Подставляя решение $\tilde{\mathbf{x}}$ обратно в (1.17), получаем:

$$\mathbf{A}\tilde{\mathbf{x}} = \tilde{\mathbf{U}}\tilde{\Sigma}\tilde{\mathbf{V}}^*\tilde{\mathbf{V}}\tilde{\Sigma}^{-1}\tilde{\mathbf{U}}^*\mathbf{b} \quad (1.21a)$$

$$\mathbf{A}\tilde{\mathbf{x}} = \tilde{\mathbf{U}}\tilde{\mathbf{U}}^*\mathbf{b}. \quad (1.21b)$$

Заметим, что $\tilde{\mathbf{U}}\tilde{\mathbf{U}}^*$ – необязательно единичная матрица, но является проекцией на пространство столбцов $\tilde{\mathbf{U}}$. Поэтому $\tilde{\mathbf{x}}$ будет точным решением системы (1.17), только если \mathbf{b} принадлежит пространству столбцов $\tilde{\mathbf{U}}$, а значит, пространству столбцов \mathbf{A} .

Вычислить псевдообратную матрицу \mathbf{A}^\dagger можно эффективно, если предварительно произвести затратное вычисление SVD. Обращение унитарных матриц $\tilde{\mathbf{U}}$ и $\tilde{\mathbf{V}}$ сводится к умножению на транспонированные матрицы, для чего требуется $O(n^2)$ операций. Обращение диагональной матрицы $\tilde{\mathbf{\Sigma}}$ еще эффективнее и требует всего $O(n)$ операций. С другой стороны, обращение плотной квадратной матрицы потребовало бы $O(n^3)$ операций.

Одномерная линейная регрессия

Регрессия – важный статистический инструмент установления связи между величинами на основе имеющихся данных [360]. Рассмотрим набор данных, изображенный на рис. 1.9. Точки, обозначенные красными крестиками, получены прибавлением гауссова белого шума к черной прямой, как показано в листинге 1.4. Мы предполагаем, что между данными имеется линейная связь, как в (1.17), и используем псевдообратную матрицу для нахождения решения с наименьшей среднеквадратической ошибкой – синей штриховой прямой с угловым коэффициентом x , – как показано в листинге 1.5.

$$\begin{bmatrix} | \\ | \\ | \end{bmatrix} \mathbf{b} = \begin{bmatrix} | \\ | \\ | \end{bmatrix} \mathbf{a} x = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^* x. \tag{1.22a}$$

$$\Rightarrow x = \tilde{\mathbf{V}}\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{U}}^* \mathbf{b}. \tag{1.22b}$$

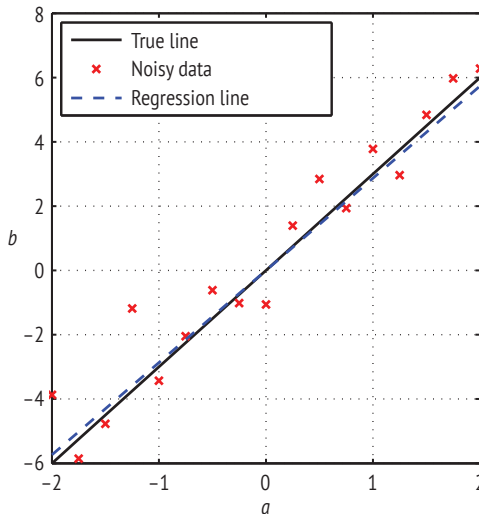


Рис. 1.9 ❖ Линейная регрессия с зашумленными данными

В (1.22b) $\tilde{\Sigma} = \|\mathbf{a}\|_2$, $\tilde{\mathbf{V}} = 1$, $\tilde{\mathbf{U}} = \mathbf{a}/\|\mathbf{a}\|_2$. Умножение на левую псевдообратную матрицу дает

$$x = \frac{\mathbf{a}^* \mathbf{b}}{\|\mathbf{a}\|_2^2}. \quad (1.23)$$

Это имеет физический смысл, если интерпретировать x как значение, которое дает наилучшее отображение нашего вектора \mathbf{a} в вектор \mathbf{b} . Такое наилучшее значение x единственно и получается в результате вычисления скалярного произведения \mathbf{b} с нормированным вектором в направлении \mathbf{a} . И мы добавляем второй нормировочный коэффициент $\|\mathbf{a}\|_2$, потому что \mathbf{a} в формуле (1.22a) не нормирован.

Отметим, что если в (1.22) взять векторы-строки вместо векторов-столбцов, то будут происходить странные вещи. Кроме того, если величина шума велика по сравнению с угловым коэффициентом x , то в точности псевдообратной матрицы произойдет фазовый переход, связанный с жесткой пороговой обработкой, описанной ниже.

Листинг 1.4 ❖ Генерирование зашумленных данных для рис. 1.9

```
x = 3; % истинный угловой коэффициент
a = [-2:.25:2]';
b = a*x + 1*randn(size(a)); % добавить шум
plot(a,x*a,'k') % истинная связь
hold on, plot(a,b,'rx') % зашумленные измерения
```

Листинг 1.5 ❖ Вычисление аппроксимации методом наименьших квадратов для рис. 1.9

```
[U,S,V] = svd(a,'econ');
xtilde = V*inv(S)*U'*b; % аппроксимации методом наименьших квадратов
plot(a,xtilde*a,'b--') % нарисовать график
```

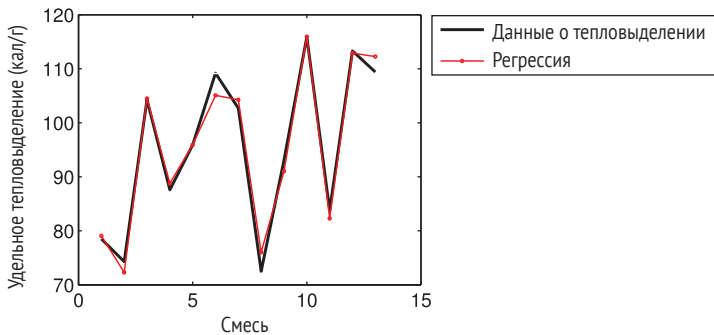


Рис. 1.10 ❖ Данные о тепловыделении для четырехкомпонентных цементных смесей

В статистике описанная выше процедура называется *линейной регрессией*. В MATLAB есть команда **regress**, а также команда **pinv**, которую тоже можно использовать.

Листинг 1.6 ❖ Альтернативные варианты метода наименьших квадратов в MATLAB

```
xtilde1 = V*inv(S)*U'*b
xtilde2 = pinv(a)*b
xtilde3 = regress(b,a)
```

Полилинейная регрессия

Пример 1: данные о тепловыделении цементных смесей

Начнем с простого встроенного в MATLAB набора данных, который описывает тепловыделение различных цементных смесей, составленных из четырех компонент. В этой задаче мы решаем систему (1.17) с $A \in \mathbb{R}^{13 \times 4}$, поскольку компонент четыре, а измерения производятся для 13 разных смесей. Наша цель – найти веса x , описывающие связь между четырьмя ингредиентами и тепловыделением. В листинге 1.7 показано, как найти решение с минимальной ошибкой. Исследуются варианты с использованием функций **regress** и **pinv**.

Листинг 1.7 ❖ Полилинейная регрессия для данных о тепловыделении цементных смесей

```
load hald; % загрузить набор данных Portland Cement
A = ingredients;
b = heat;

[U,S,V] = svd(A,'econ');
x = V*inv(S)*U'*b; % решить систему Ax=b методом SVD

plot(b,'k'); hold on % построить график
plot(A*x,'r-o',); % аппроксимация графика

x = regress(b,A); % вариант 1 (regress)
x = pinv(A)*b; % вариант 2 (pinv)
```

Пример 2: данные о недвижимости в Бостоне

В этом примере мы исследуем более крупный набор данных, чтобы определить, какие факторы лучше всего предсказывают цены на бостонском рынке недвижимости [234]. Этот набор можно скачать из репозитория машинного обучения UCI Machine Learning Repository [24].

Существует 13 атрибутов, коррелирующих с ценой дома, например: индекс преступности на душу населения и ставка налога на имущество. Мы построили регрессионную модель влияния этих факторов на цену и на рис. 1.11 на-

рисовали график, на котором лучшее предсказание цены сопоставляется с истинной стоимостью дома. На рис. 1.12 показаны коэффициенты регрессии. Хотя стоимость дома предсказана неточно, тренды хорошо согласуются. Часто бывает, что простая линейная аппроксимация плохо улавливает выбросы с наибольшими значениями; такую ситуацию мы наблюдаем и в этом примере.

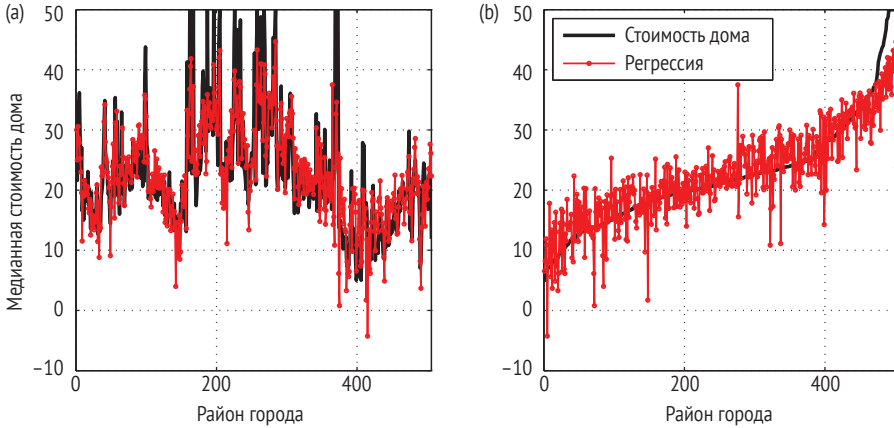


Рис. 1.11 ❖ Полилинейная регрессия цен домов с различными факторами: (a) неотсортированные данные; (b) данные отсортированы по стоимости дома

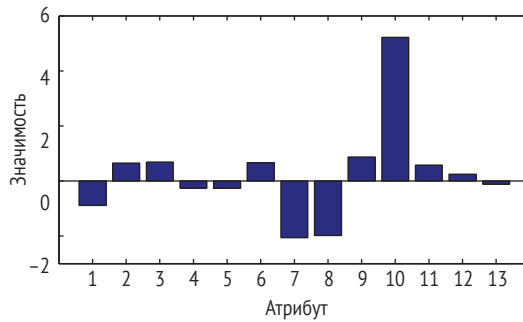


Рис. 1.12 ❖ Значимость различных атрибутов в регрессии

Этот набор данных содержит цены и атрибуты для 506 домов, так что матрица атрибутов имеет размер 506×13 . Важно дополнить эту матрицу столбцом, состоящим из единиц, чтобы учесть возможность ненулевого постоянного смещения в формуле регрессии. В случае одномерной регрессии это соответствует свободному члену.

Листинг 1.8 ❖ Полилинейная регрессия для данных о стоимости домов в Бостоне

```
load housing.data
```

```
b = housing(:,14); % стоимости домов в тысячах долларов
A = housing(:,1:13); % прочие факторы
```

```
A = [A ones(size(A,1),1)]; % дополнить единичными свободными членами
x = regress(b,A);
plot(b,'k-o');
hold on, plot(A*x,'r-o');

[b sortind] = sort(housing(:,14)); % отсортированные стоимости
plot(b,'k-o')
hold on, plot(A(sortind,:)*x,'r-o')
```

Предостережение

Вообще говоря, матрица U , столбцами которой являются левые сингулярные векторы X , – унитарная квадратная матрица. Поэтому $U^*U = UU^* = I_{n \times n}$. Однако чтобы вычислить псевдообратную матрицу для X , мы должны вычислить $X^\dagger = \tilde{V}\tilde{\Sigma}^{-1}\tilde{U}^*$, потому что только $\tilde{\Sigma}$ обратима (если все сингулярные значения отличны от нуля), тогда как Σ в общем случае необратима (она даже в общем случае не является квадратной).

До сих пор мы предполагали, что $X = \tilde{U}\tilde{\Sigma}\tilde{V}^*$ – точное SVD, так что ранг r включает все ненулевые сингулярные значения. Это гарантирует, что матрица $\tilde{\Sigma}$ обратима.

Трудности начинаются при работе с усеченным базисом, образованным левыми сингулярными векторами \tilde{U} . По-прежнему верно, что $\tilde{U}^*\tilde{U} = I_{r \times r}$, где r – ранг X . Однако $\tilde{U}\tilde{U}^* \neq I_{n \times n}$, что легко проверить численно на простом примере. Предположение о том, что $\tilde{U}\tilde{U}^*$ равно единичной матрице, – одно из самых типичных заблуждений при использовании SVD¹.

```
>> tol = 1.e-16;
>> [U,S,V] = svd(X,'econ')
>> r = max(find(diag(S)>max(S(:))*tol));
>> invX = V(:,1:r)*S(1:r,1:r)*U(:,1:r)'; % только приближенно
```

1.5. МЕТОД ГЛАВНЫХ КОМПОНЕНТ (PCA)

Метод главных компонент (PCA) – одно из основных применений SVD, он позволяет, ориентируясь только на данные, построить систему координат для представления коррелированных данных высокой размерности. При этом используются корреляционные матрицы, описанные в разделе 1.3. Важно, что в методе PCA, перед тем как применять SVD, производится предварительная обработка данных: вычитание среднего и приведение к единичной дисперсии. Геометрия результирующей системы координат определяется главными компонентами (principal component – PC), которые не коррелируют между собой (ортогональны), но имеют максимальную корреляцию с результатами измерений. Теория была разработана в 1901 году Пирсоном

¹ Его не избежали и авторы, по ошибке включив это неверное тождество в ранний вариант работы [96].

[418] и независимо в 1930-х годах Хотеллингом [256, 257]. Работа Jolliffe [268] содержит хорошее справочное пособие.

Обычно в одном эксперименте производится несколько измерений, которые образуют вектор-строку. Эти измерения могут быть свойствами наблюдаемой величины, например демографическими признаками одного человека. Выполняется несколько экспериментов, и все векторы-строки собираются в большую матрицу \mathbf{X} . Если говорить о демографии, то экспериментальные данные могут быть собраны путем опроса. Заметим, что такое соглашение – \mathbf{X} состоит из строк признаков – отличается от соглашения, принятого в других частях этой главы, где отдельные «снимки» признаков расположены по столбцам. Но мы решили в этом разделе не вступать в противоречие с литературой по PCA. Матрица по-прежнему имеет размер $n \times m$ и может содержать больше строк, чем столбцов, или наоборот.

Вычисление

Теперь вычислим среднее по строкам $\bar{\mathbf{x}}$ (т. е. среднее всех строк) и вычтем его из \mathbf{X} . Среднее $\bar{\mathbf{x}}$ определяется по формуле

$$\bar{\mathbf{x}}_j = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_{ij}, \quad (1.24)$$

а средняя матрица – по формуле

$$\bar{\mathbf{X}} = \begin{bmatrix} 1 \\ \vdots \\ \bar{\mathbf{x}} \\ 1 \end{bmatrix}. \quad (1.25)$$

Вычитая $\bar{\mathbf{X}}$ из \mathbf{X} , получаем матрицу \mathbf{V} с нулевым средним:

$$\mathbf{V} = \mathbf{X} - \bar{\mathbf{X}} \quad (1.26)$$

Ковариационная матрица строк \mathbf{V} определяется формулой

$$\mathbf{C} = \frac{1}{n-1} \mathbf{V}^* \mathbf{V}. \quad (1.27)$$

Первая главная компонента \mathbf{u}_1 определяется так:

$$\mathbf{u}_1 = \operatorname{argmax}_{\|\mathbf{u}_1\|=1} \mathbf{u}_1^* \mathbf{V}^* \mathbf{V} \mathbf{u}_1. \quad (1.28)$$

Это собственный вектор $\mathbf{V}^* \mathbf{V}$, соответствующий наибольшему собственному значению. Теперь понятно, что \mathbf{u}_1 – левый сингулярный вектор \mathbf{V} , соответствующий наибольшему сингулярному значению.

Главные компоненты можно получить, вычислив спектральное разложение \mathbf{C} :

$$\mathbf{C} \mathbf{V} = \mathbf{V} \mathbf{D}, \quad (1.29)$$

которое гарантированно существует, потому что матрица \mathbf{C} эрмитова.

Команда `pca`

В MATLAB имеются дополнительные команды `pca` и `princomp` (основанная на `pca`) для метода главных компонент:

```
>> [V,score,s2] = pca(X);
```

Матрица \mathbf{V} эквивалентна матрице \mathbf{V} из сингулярного разложения \mathbf{X} с точностью до изменения знака столбцов. Вектор $\mathbf{s2}$ содержит собственные значения ковариационной матрицы \mathbf{X} , которые называются также дисперсиями главных компонент; это квадраты сингулярных значений. Переменная `score` содержит координаты каждой строки \mathbf{V} (данные после вычитания среднего) в направлениях главных компонент. Вообще говоря, мы часто предпочитаем использовать команду `svd` в сочетании с различными шагами постобработки, описанными ранее в этом разделе.

Пример: данные с гауссовым шумом

Рассмотрим зашумленное облако данных на рис. 1.13 (а), построенном программой в листинге 1.9. Данные генерируются путем выборки 10 000 векторов из двумерного нормального распределения с нулевым средним и единичной дисперсией. Эти векторы затем масштабируются в направлениях x и y с коэффициентами в табл. 1.1 и поворачиваются на угол $\pi/3$. Наконец, все облако данных подвергается параллельному переносу, так что его центр располагается в точке $\mathbf{x}_C = [2 \ 1]^T$.

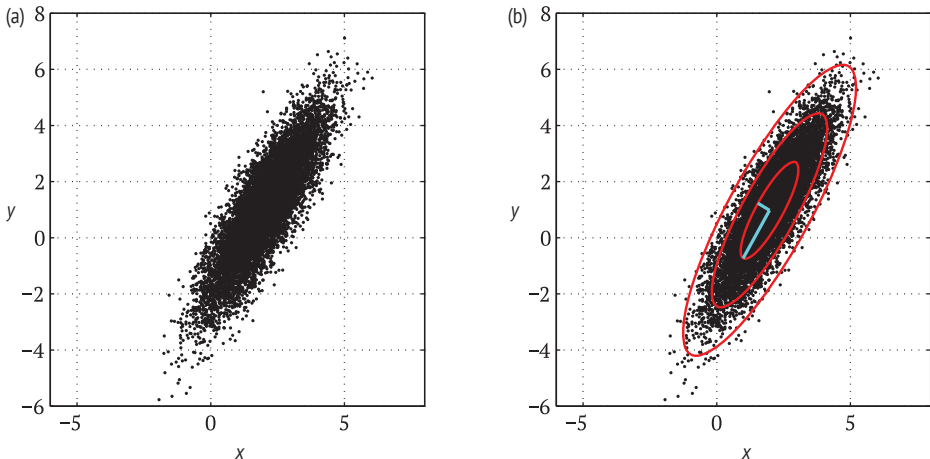


Рис. 1.13 ❖ Главные компоненты улавливают дисперсию гауссовых данных после вычитания среднего (а). На рисунке (б) показаны эллипсы для первых трех стандартных отклонений (красным цветом) и два левых сингулярных вектора, умноженных на сингулярные значения ($\sigma_1 \mathbf{u}_1 + \mathbf{x}_C$ и $\sigma_2 \mathbf{u}_2 + \mathbf{x}_C$, голубым цветом)

Таблица 1.1. Стандартное отклонение данных и нормированные сингулярные значения

	σ_1	σ_2
Данные	2	0.5
SVD	1.974	0.503

Программа в листинге 1.10 выполняет PCA и рисует на графике доверительные интервалы при нескольких стандартных отклонениях, как показано на рис. 1.13 (b). В табл. 1.1 также приведены сингулярные значения, соответствующие коэффициентам масштабирования. Матрица \mathbf{U} из сингулярного разложения очень близка к матрице поворота с точностью до знака чисел в одном столбце:

$$\mathbf{R}_{\pi/3} = \begin{bmatrix} 0.5 & -0.8660 \\ 0.8660 & 0.5 \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} -0.4998 & -0.8662 \\ -0.8662 & 0.4998 \end{bmatrix}.$$

Листинг 1.9 ❖ Генерирование зашумленного облака данных для иллюстрации PCA

```
xC = [2; 1;]; % центр данных (среднее)
sig = [2; .5;]; % главные оси

theta = pi/3; % повернуть облако на угол pi/3
R = [cos(theta) -sin(theta); % матрица поворота
sin(theta) cos(theta)];

nPoints = 10000; % создать 10 000 точек
X = R*diag(sig)*randn(2,nPoints) + diag(xC)*ones(2,nPoints);
scatter(X(1,:),X(2,:), 'k.', 'LineWidth', 2)
```

Листинг 1.10 ❖ Вычислить главные компоненты и нарисовать доверительные интервалы

```
Xavg = mean(X,2); % вычислить среднее
B = X - Xavg*ones(1,nPoints); % данные после вычитания среднего
[U,S,V] = svd(B/sqrt(nPoints), 'econ'); % PCA методом SVD
scatter(X(1,:),X(2,:), 'k.', 'LineWidth', 2) % нанести данные на график

theta = (0:.01:1)*2*pi;
Xstd = U*S*[cos(theta); sin(theta)]; % доверит. инт. для 1-го стд. откл.

plot(Xavg(1)+Xstd(1,:),Xavg(2) + Xstd(2,:), 'r-')
plot(Xavg(1)+2*Xstd(1,:),Xavg(2) + 2*Xstd(2,:), 'r-')
plot(Xavg(1)+3*Xstd(1,:),Xavg(2) + 3*Xstd(2,:), 'r-')
```

Наконец, главные компоненты можно вычислить также командой **pca**:

```
>> [V,score,s2] = pca(X);
>> norm(V*score' - B)
```

```
ans =
    2.2878e-13
```


Список литературы

- [1] Ralph Abraham, Jerrold E. Marsden, and Tudor Ratiu. *Manifolds, Tensor Analysis, and Applications*, volume 75 of *Applied Mathematical Sciences*. Springer-Verlag, 1988.
- [2] Ralph Abraham and Jerrold E. Marsden. *Foundations of Mechanics*, volume 36. Benjamin/Cummings Publishing Company Reading, Massachusetts, 1978.
- [3] Mradul Agrawal, Sandeep Vidyashankar, and Ke Huang. On-chip implementation of ECoG signal data decoding in brain-computer interface. In *Mixed-Signal Testing Workshop (IMSTW), 2016 IEEE 21st International*, pages 1–6. IEEE, 2016.
- [4] Rakesh Agrawal, Ramakrishnan Srikant, et al. Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB*, volume 1215, pages 487–499, 1994.
- [5] Hyo-Sung Ahn, YangQuan Chen, and Kevin L. Moore. Iterative learning control: Brief survey and categorization. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(6):1099–1121, 2007.
- [6] Hirotugu Akaike. Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*, 21(1):243–247, 1969.
- [7] Hirotugu Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723, 1974.
- [8] W. O. Amrein and Anne-Marie Berthier. On support properties of Lp-functions and their Fourier transforms. *Journal of Functional Analysis*, 24(3):258–267, 1977.
- [9] David Amsallem, Julien Cortial, and Charbel Farhat. On-demand cfd-based aeroelastic predictions using a database of reduced-order bases and models. In *47th AIAA Aerospace Sciences Meeting Including The New Horizons Forum and Aerospace Exposition*, page 800, 2009.
- [10] David Amsallem and Charbel Farhat. An online method for interpolating linear parametric reduced-order models. *SIAM Journal on Scientific Computing*, 33(5):2169–2198, 2011.
- [11] David Amsallem, Matthew J. Zahr, and Kyle Washabaugh. Fast local reduced basis updates for the efficient reduction of nonlinear systems with hyper-reduction. *Advances in Computational Mathematics*, 41(5):1187–1230, 2015.
- [12] Joakim Andén and Stéphane Mallat. Deep scattering spectrum. *IEEE Transactions on Signal Processing*, 62(16):4114–4128, 2014.
- [13] Edward Anderson, Zhaojun Bai, Christian Bischof, Susan Blackford, James Demmel, Jack Dongarra, Jeremy Du Croz, Anne Greenbaum, S. Hammerling, Alan McKenney, et al. *LAPACK Users' Guide*, volume 9. SIAM, 1999.
- [14] Jeffrey L. Anderson. An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, 129(12):2884–2903, 2001.
- [15] Claus A. Andersson and Rasmus Bro. The n-way toolbox for matlab. *Chemo-metrics and Intelligent Laboratory Systems*, 52(1):1–4, 2000.

- [16] Marc Antonini, Michel Barlaud, Pierre Mathieu, and Ingrid Daubechies. Image coding using wavelet transform. *IEEE Transactions on Image Processing*, 1(2):205–220, 1992.
- [17] Athanasios C. Antoulas. *Approximation of large-scale dynamical systems*. SIAM, 2005.
- [18] Hassan Arbabi and Igor Mezic'. Ergodic theory, dynamic mode decomposition and computation of spectral properties of the Koopman operator. *SIAM Journal on Applied Dynamical Systems*, 16 (4):2096–2126, 2017.
- [19] Kartik B. Ariyur and Miroslav Krstic'. *Real-Time Optimization by Extremum-Seeking Control*. Wiley, Hoboken, New Jersey, 2003.
- [20] Travis Askham and J. Nathan Kutz. Variable projection methods for an optimized dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 17 (1):380–416, 2018.
- [21] Patricia Astrid. Fast reduced order modeling technique for large scale LTV systems. In *American Control Conference, 2004. Proceedings of the 2004*, volume 1, pages 762–767. IEEE, 2004.
- [22] Karl Johan Aström and Richard M. Murray. *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton University Press, 2010.
- [23] M. F. A. Azeez and A. F. Vakakis. Proper orthogonal decomposition (POD) of a class of vibroimpact oscillations. *Journal of Sound and Vibration*, 240(5): 859–889, 2001.
- [24] K. Bache and M. Lichman. UCI machine learning repository, 2013.
- [25] Brett W. Bader and Tamara G. Kolda. Efficient MATLAB computations with sparse and factored tensors. *SIAM Journal on Scientific Computing*, 30(1):205–231, December 2007.
- [26] S. Bagheri, L. Brandt, and D. S. Henningson. Input-output analysis, model reduction and control of the flat-plate boundary layer. *Journal of Fluid Mechanics*, 620:263–298, 2009.
- [27] S. Bagheri, J. Hoepffner, P. J. Schmid, and D. S. Henningson. Input-output analysis and control design applied to a linear model of spatially developing flows. *Applied Mechanics Reviews*, 62(2):020803–1..27, 2009.
- [28] Shervin Bagheri. Koopman-mode decomposition of the cylinder wake. *Journal of Fluid Mechanics*, 726:596–623, 2013.
- [29] Z. Bai, S. L. Brunton, B. W. Brunton, J. N. Kutz, E. Kaiser, A. Spohn, and B. R. Noack. Data- driven methods in fluid dynamics: Sparse classification from experimental data. In *Invited Chapter for Whither Turbulence and Big Data in the 21st Century*, 2015.
- [30] Z. Bai, E. Kaiser, J. L. Proctor, J. N. Kutz, and S. L. Brunton. Dynamic mode decomposition for compressive system identification. *arXiv preprint arXiv:1710.07737*, 2017.
- [31] Zhe Bai, Thakshila Wimalajeewa, Zachary Berger, Guannan Wang, Mark Glauser, and Pramod K. Varshney. Low-dimensional approach for reconstruction of airfoil data via compressive sensing. *AIAA Journal*, 53(4):920–933, 2014.
- [32] Maciej J. Balajewicz, Earl H. Dowell, and Bernd R. Noack. Low-dimensional modelling of high-Reynolds-number shear flows incorporating constraints from the Navier–Stokes equation. *Journal of Fluid Mechanics*, 729:285–308, 2013.

- [33] Madhusudhanan Balasubramanian, Stanislav Zabic, Christopher Bowd, Hilary W. Thompson, Peter Wolenski, S. Sitharama Iyengar, Bijaya B. Karki, and Linda M. Zangwill. A framework for detecting glaucomatous progression in the optic nerve head of an eye using proper orthogonal decomposition. *IEEE Transactions on Information Technology in Biomedicine*, 13(5):781–793, 2009.
- [34] B. Bamieh and L. Giarré. Identification of linear parameter varying models. *International Journal of Robust and Nonlinear Control*, 12:841–853, 2002.
- [35] Andrzej Banaszuk, Kartik B. Ariyur, Miroslav Krstic', and Clas A. Jacobson. An adaptive algorithm for control of combustion instability. *Automatica*, 40(11):1965–1972, 2004.
- [36] Andrzej Banaszuk, Satish Narayanan, and Youping Zhang. Adaptive control of flow separation in a planar diffuser. *AIAA paper*, 617:2003, 2003.
- [37] Andrzej Banaszuk, Youping Zhang, and Clas A. Jacobson. Adaptive control of combustion instability using extremum-seeking. In *American Control Conference, 2000. Proceedings of the 2000*, volume 1, pages 416–422. IEEE, 2000.
- [38] S. P. Banks. Infinite-dimensional Carleman linearization, the Lie series and optimal control of non-linear partial differential equations. *International Journal of Systems Science*, 23(5):663–675, 1992.
- [39] R. G. Baraniuk. Compressive sensing. *IEEE Signal Processing Magazine*, 24(4):118–120, 2007.
- [40] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Transactions on Information Theory*, 56(4):1982–2001, 2010.
- [41] Maxime Barrault, Yvon Maday, Ngoc Cuong Nguyen, and Anthony T. Patera. An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathématique*, 339(9):667–672, 2004.
- [42] J. Basley, L. R. Pastur, N. Delprat, and F. Lusseyran. Space-time aspects of a three-dimensional multi-modulated open cavity flow. *Physics of Fluids (1994-present)*, 25(6):064105, 2013.
- [43] J. Basley, L. R. Pastur, F. Lusseyran, T. M. Faure, and N. Delprat. Experimental investigation of global structures in an incompressible cavity flow using time-resolved PIV. *Experiments in Fluids*, 50(4):905–918, 2011.
- [44] Walter Baur and Volker Strassen. The complexity of partial derivatives. *Theoretical Computer Science*, 22(3):317–330, 1983.
- [45] J. F. Beaudoin, O. Cadot, J. L. Aider, and J. E. Wesfreid. Bluff-body drag reduction by extremum-seeking control. *Journal of Fluids and Structures*, 22:973–978, 2006.
- [46] Jean-Francois Beaudoin, Olivier Cadot, Jean-Luc Aider, and José-Eduardo Wesfreid. Drag reduction of a bluff body using adaptive control methods. *Physics of Fluids*, 18(8):085107, 2006.
- [47] R. Becker, R. King, R. Petz, and W. Nitsche. Adaptive closed-loop control on a high-lift configuration using extremum seeking. *AIAA Journal*, 45(6):1382–92, 2007.
- [48] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 19(7):711–720, 1997.

- [49] G. Bellani. Experimental studies of complex flows through image-based techniques. 2011.
- [50] Brandt A. Belson, Jonathan H. Tu, and Clarence W. Rowley. Algorithm 945: modred – a parallelized model reduction library. *ACM Transactions on Mathematical Software*, 40(4):30, 2014.
- [51] Michael Benedicks. On Fourier transforms of functions supported on sets of finite Lebesgue measure. *Journal of Mathematical Analysis and Applications*, 106(1):180–183, 1985.
- [52] Yoshua Bengio, Pascal Lamblin, Dan Popovici, and Hugo Larochelle. Greedy layer-wise training of deep networks. In *Advances in Neural Information Processing Systems*, pages 153–160, 2007.
- [53] P. Benner, S. Gugercin, and K Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 57(4):483–531, 2015.
- [54] Peter Benner, Albert Cohen, Mario Ohlberger, and Karen Willcox. *Model Reduction and Approximation: Theory and Algorithms*, volume 15. SIAM, 2017.
- [55] Peter Benner, Jing-Rebecca Li, and Thilo Penzl. Numerical solution of large-scale Lyapunov equations, Riccati equations, and linear-quadratic optimal control problems. *Numerical Linear Algebra with Applications*, 15(9):755–777, 2008.
- [56] E. Berger, M. Sastuba, D. Vogt, B. Jung, and H. B. Amor. Estimation of perturbations in robotic behavior using dynamic mode decomposition. *Journal of Advanced Robotics*, 29(5):331–343, 2015.
- [57] G. Berkooz, P. Holmes, and J. L. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual Review of Fluid Mechanics*, 25:539–575, 1993.
- [58] Gregory Beylkin, Ronald Coifman, and Vladimir Rokhlin. Fast wavelet transforms and numerical algorithms i. *Communications on Pure and Applied Mathematics*, 44(2):141–183, 1991.
- [59] Stephen A. Billings. *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*. John Wiley & Sons, 2013.
- [60] P. Binetti, K. B. Ariyur, M. Krstić, and F. Bernelli. Formation flight optimization using extremum seeking feedback. *Journal of Guidance, Control, and Dynamics*, 26(1):132–142, 2003.
- [61] G. D. Birkhoff and B. O. Koopman. Recent contributions to the ergodic theory. *Proceedings of the National Academy of Sciences*, 18(3):279–282, 1932.
- [62] George D. Birkhoff. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences*, 17(12):656–660, 1931.
- [63] Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford university press, 1995.
- [64] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer New York, 2006.
- [65] D. A. Bistrrian and I. M. Navon. An improved algorithm for the shallow water equations model reduction: Dynamic mode decomposition vs POD. *International Journal for Numerical Methods in Fluids*, 2015.
- [66] D. A. Bistrrian and I. M. Navon. Randomized dynamic mode decomposition for non-intrusive reduced order modelling. *International Journal for Numerical Methods in Engineering*, 2016.

- [67] Paola Bondi, Giuseppe Casalino, and Lucia Gambardella. On the iterative learning control theory for robotic manipulators. *IEEE Journal on Robotics and Automation*, 4(1):14–22, 1988.
- [68] Josh Bongard and Hod Lipson. Automated reverse engineering of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 104(24):9943–9948, 2007.
- [69] Jorge Luis Borges. The library of Babel. *Collected Fictions*, 1998.
- [70] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pages 144–152. ACM, 1992.
- [71] H. Boulard and Y. Kamp. Autoassociative memory by multilayer perceptron and singular values decomposition. *Biol Cybern*, 59:291–294, 1989.
- [72] George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 2015.
- [73] Stephen Boyd, Leon O. Chua, and Charles A. Desoer. Analytical foundations of Volterra series. *IMA Journal of Mathematical Control and Information*, 1(3):243–282, 1984.
- [74] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2009.
- [75] Antoneta Iuliana Bratcu, Iulian Munteanu, Seddik Bacha, and Bertrand Raison. Maximum power point tracking of grid-connected photovoltaic arrays by using extremum seeking control. *CEAI*, 10(4):3–12, 2008.
- [76] Leo Breiman. Better subset regression using the nonnegative garrote. *Technometrics*, 37(4):373–384, 1995.
- [77] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- [78] Leo Breiman et al. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3):199–231, 2001.
- [79] Leo Breiman, Jerome Friedman, Charles J. Stone, and Richard A. Olshen. *Classification and Regression Trees*. CRC press, 1984.
- [80] I. Bright, G. Lin, and J. N. Kutz. Compressive sensing and machine learning strategies for characterizing the flow around a cylinder with limited pressure measurements. *Physics of Fluids*, 25(127102):1–15, 2013.
- [81] Ido Bright, Guang Lin, and J. Nathan Kutz. Classification of spatio-temporal data via asynchronous sparse sampling: Application to flow around a cylinder. *Multiscale Modeling & Simulation*, 14 (2), 823–838, 2016.
- [82] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117, 1998.
- [83] Douglas Bristow, Marina Tharayil, Andrew G. Alleyne, et al. A survey of iterative learning control. *Control Systems, IEEE*, 26(3):96–114, 2006.
- [84] Rasmus Bro. Parafac. tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, 38(2):149–171, 1997.
- [85] Alexander Broad, T. D. Murphey, and Brenna Argall. Learning models for shared control of human-machine systems with unknown dynamics. *Robotics: Science and Systems Proceedings*, 2017.
- [86] Roger W. Brockett. Volterra series and geometric control theory. *Automatica*, 12(2):167–176, 1976.

- [87] David S. Broomhead and David Lowe. Radial basis functions, multi-variable functional interpolation and adaptive networks. Technical report, Royal Signals and Radar Establishment Malvern (United Kingdom), 1988.
- [88] D. S. Broomhead and R Jones. Time-series analysis. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 423, pages 103–121. The Royal Society, 1989.
- [89] B. W. Brunton, S. L. Brunton, J. L. Proctor, and J. N. Kutz. Sparse sensor placement optimization for classification. *SIAM Journal on Applied Mathematics*, 76(5):2099–2122, 2016.
- [90] B. W. Brunton, L. A. Johnson, J. G. Ojemann, and J. N. Kutz. Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *Journal of Neuroscience Methods*, 258:1–15, 2016.
- [91] S. L. Brunton, B. W. Brunton, J. L. Proctor, E. Kaiser, and J. N. Kutz. Chaos as an intermittently forced linear system. *Nature Communications*, 8(19):1–9, 2017.
- [92] S. L. Brunton, B. W. Brunton, J. L. Proctor, and J. N. Kutz. Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control. *PLoS ONE*, 11(2):e0150171, 2016.
- [93] S. L. Brunton, X. Fu, and J. N. Kutz. Extremum-seeking control of a mode-locked laser. *IEEE Journal of Quantum Electronics*, 49(10):852–861, 2013.
- [94] S. L. Brunton and B. R. Noack. Closed-loop turbulence control: Progress and challenges. *Applied Mechanics Reviews*, 67:050801–1–050801–48, 2015.
- [95] S. L. Brunton, J. L. Proctor, and J. N. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.
- [96] S. L. Brunton, J. L. Proctor, J. H. Tu, and J. N. Kutz. Compressed sensing and dynamic mode decomposition. *Journal of Computational Dynamics*, 2(2):165–191, 2015.
- [97] S. L. Brunton and C. W. Rowley. Maximum power point tracking for photovoltaic optimization using ripple-based extremum seeking control. *IEEE Transactions on Power Electronics*, 25(10):2531–2540, 2010.
- [98] S. L. Brunton, J. H. Tu, I. Bright, and J. N. Kutz. Compressive sensing and low-rank libraries for classification of bifurcation regimes in nonlinear dynamical systems. *SIAM Journal on Applied Dynamical Systems*, 13(4):1716–1732, 2014.
- [99] Steven L. Brunton, Xing Fu, and J. Nathan Kutz. Self-tuning fiber lasers. *IEEE Journal of Selected Topics in Quantum Electronics*, 20(5), 2014.
- [100] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Sparse identification of nonlinear dynamics with control (SINDYc). *IFAC NOLCOS*, 49(18):710–715, 2016.
- [101] D. Buche, Peter Stoll, Rolf Dornberger, and Petros Koumoutsakos. Multi-objective evolutionary algorithm for the optimization of noisy combustion processes. *Systems, Man, and Cybernetics, Part C: Applications and Reviews*, *IEEE Transactions on*, 32(4):460–473, 2002.
- [102] Marko Budišić and Igor Mezic. An approximate parametrization of the ergodic partition using time averaged observables. In *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pages 3162–3168. IEEE, 2009.

- [103] Marko Budišić and Igor Mezic. Geometry of the ergodic quotient reveals coherent structures in flows. *Physica D: Nonlinear Phenomena*, 241(15):1255–1269, 2012.
- [104] Marko Budišić, Ryan Mohr, and Igor Mezic. Applied Koopmanism a). *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(4):047510, 2012.
- [105] Kenneth P. Burnham and David R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer Science & Business Media, 2003.
- [106] Peter A. Businger and Gene H. Golub. Algorithm 358: Singular value decomposition of a complex matrix [f1, 4, 5]. *Communications of the ACM*, 12(10):564–565, 1969.
- [107] Eduardo F. Camacho and Carlos Bordons Alba. *Model Predictive Control*. Springer Science & Business Media, 2013.
- [108] Erik Cambria, Guang-Bin Huang, Liyanaarachchi Lekamalage Chamara Kasun, Hongming Zhou, Chi Man Vong, Jiarun Lin, Jianping Yin, Zhiping Cai, Qiang Liu, Kuan Li, et al. Extreme learning machines [trends & controversies]. *IEEE Intelligent Systems*, 28(6):30–59, 2013.
- [109] E. J. Candès. Compressive sensing. *Proceedings of the International Congress of Mathematics*, 2006.
- [110] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):11–1–11–37, 2011.
- [111] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.
- [112] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications in Pure and Applied Mathematics*, 8(1207–1223), 59.
- [113] E. J. Candès and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006.
- [114] E. J. Candès and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, pages 21–30, 2008.
- [115] Emmanuel J. Candes and Terence Tao. Decoding by linear programming. *Information Theory, IEEE Transactions on*, 51(12):4203–4215, 2005.
- [116] Yanhua Cao, Jiang Zhu, Zhendong Luo, and IM Navon. Reduced-order modeling of the upper tropical pacific ocean model using proper orthogonal decomposition. *Computers & Mathematics with Applications*, 52(8):1373–1386, 2006.
- [117] Yanhua Cao, Jiang Zhu, I Michael Navon, and Zhendong Luo. A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. *International Journal for Numerical Methods in Fluids*, 53(10):1571–1583, 2007.
- [118] Kevin Carlberg, Matthew Barone, and Harbir Antil. Galerkin v. least-squares Petrov–Galerkin projection in nonlinear model reduction. *Journal of Computational Physics*, 330:693–734, 2017.
- [119] Kevin Carlberg, Charbel Bou-Mosleh, and Charbel Farhat. Efficient nonlinear model reduction via a least-squares Petrov–Galerkin projection and

- compressive tensor approximations. *International Journal for Numerical Methods in Engineering*, 86(2):155–181, 2011.
- [120] Kevin Carlberg, Charbel Farhat, Julien Cortial, and David Amsallem. The GNAT method for nonlinear model reduction: effective implementation and application to computational fluid dynamics and turbulent flows. *Journal of Computational Physics*, 242:623–647, 2013.
- [121] Torsten Carleman. Application de la théorie des équations intégrales linéaires aux systèmes d'équations différentielles non linéaires. *Acta Mathematica*, 59(1):63–87, 1932.
- [122] Torsten Carleman. Sur la théorie de l'équation intégrodifférentielle de boltzmann. *Acta Mathematica*, 60(1):91–146, 1933.
- [123] Torsten Carleman. Sur les systemes lineaires aux dérivées partielles du premier ordrea deux variables. *CR Acad. Sci. Paris*, 197:471–474, 1933.
- [124] J. Douglas Carroll and Jih-Jie Chang. Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition. *Psychometrika*, 35:283–319, 1970.
- [125] Rick Chartrand. Numerical differentiation of noisy, nonsmooth data. *ISRN Applied Mathematics*, 2011, 2011.
- [126] Anindya Chatterjee. An introduction to the proper orthogonal decomposition. *Current Science*, 78(7):808–817, 2000.
- [127] S. Chaturantabut and D. C. Sorensen. Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing*, 32(5):2737–2764, 2010.
- [128] K. K. Chen and C. W. Rowley. Normalized coprime robust stability and performance guarantees for reduced-order controllers. *IEEE Transactions on Automatic Control*, 58(4):1068–1073, 2013.
- [129] K. K. Chen, J. H. Tu, and C. W. Rowley. Variants of dynamic mode decomposition: Boundary condition, Koopman, and Fourier analyses. *Journal of Nonlinear Science*, 22(6):887–915, 2012.
- [130] Yangquan Chen, Kevin L Moore, and Hyo-Sung Ahn. Iterative learning control. In *Encyclopedia of the Sciences of Learning*, pages 1648–1652. Springer, 2012.
- [131] Steve Cherry. Singular value decomposition analysis and canonical correlation analysis. *Journal of Climate*, 9(9):2003–2009, 1996.
- [132] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder- decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.
- [133] J. Y. Choi, M. Krstić, K. B. Ariyur, and J. S. Lee. Extremum seeking control for discrete-time systems. *IEEE Transactions on Automatic Control*, 47(2):318–323, FEB 2002.
- [134] Youngsoo Choi, David Amsallem, and Charbel Farhat. Gradient-based constrained optimization using a database of linear reduced-order models. *arXiv preprint arXiv:1506.07849*, 2015.
- [135] T. Colonius and K. Taira. A fast immersed boundary method using a nullspace approach and multi-domain far-field boundary conditions. *Computer Methods in Applied Mechanics and Engineering*, 197:2131–2146, 2008.

- [136] James W. Cooley, Peter A. W. Lewis, and Peter D. Welch. Historical notes on the fast Fourier transform. *Proceedings of the IEEE*, 55(10):1675–1677, 1967.
- [137] James W. Cooley and John W Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19(90):297–301, 1965.
- [138] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.
- [139] Mark C. Cross and Pierre C. Hohenberg. Pattern formation outside of equilibrium. *Reviews of Modern Physics*, 65(3):851, 1993.
- [140] James P. Crutchfield and Bruce S. McNamara. Equations of motion from a data series. *Complex Systems*, 1:417–452, 1987.
- [141] Magnus Dam, Morten Brøns, Jens Juul Rasmussen, Volker Naulin, and Jan S. Hesthaven. Sparse identification of a predator-prey system from simulation data of a convection model. *Physics of Plasmas*, 24(2):022310, 2017.
- [142] Bryan C. Daniels and Ilya Nemenman. Automated adaptive inference of phenomenological dynamical models. *Nature Communications*, 6, 2015.
- [143] Bryan C. Daniels and Ilya Nemenman. Efficient inference of parsimonious phenomenological models of cellular dynamics using s-systems and alternating regression. *PloS one*, 10(3):e0119821, 2015.
- [144] Suddhasattwa Das and Dimitrios Giannakis. Delay-coordinate maps and the spectra of Koopman operators. *arXiv preprint arXiv:1706.08544*, 2017.
- [145] Ingrid Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE transactions on information theory*, 36(5):961–1005, 1990.
- [146] Lawrence Davis et al. *Handbook of Genetic Algorithms*, volume 115. Van Nostrand Reinhold New York, 1991.
- [147] Scott T. M. Dawson, Maziar S. Hemati, Matthew O. Williams, and Clarence W. Rowley. Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Experiments in Fluids*, 57(3):1–19, 2016.
- [148] Arthur P. Dempster, Nan M. Laird, and Donald B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (methodological)*, pages 1–38, 1977.
- [149] Santosh Devasia, Degang Chen, and Brad Paden. Nonlinear inversion-based output tracking. *Automatic Control, IEEE Transactions on*, 41(7):930–942, 1996.
- [150] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.
- [151] D. L. Donoho and M. Gavish. Code supplement to “The optimal hard threshold for singular values is $4/\sqrt{3}$ ”. <http://purl.stanford.edu/vg705qn9070>, 2014.
- [152] David Donoho. 50 years of data science. In *Based on a Presentation at the Tukey Centennial Workshop*. NJ Princeton, 2015.
- [153] David L. Donoho, Iain M Johnstone, Jeffrey C Hoch, and Alan S Stern. Maximum entropy and the nearly black object. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 41–81, 1992.
- [154] David L. Donoho and Jain M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.
- [155] J. C. Doyle. Guaranteed margins for LQG regulators. *IEEE Transactions on Automatic Control*, 23(4):756–757, 1978.

- [156] J. C. Doyle, K. Glover, P. P. Khargonekar, and B. A. Francis. State-space solutions to standard H_2 and H_∞ control problems. *IEEE Transactions on Automatic Control*, 34(8):831–847, 1989.
- [157] John C. Doyle, Bruce A. Francis, and Allen R. Tannenbaum. *Feedback Control Theory*. Courier Corporation, 2013.
- [158] Petros Drineas and Michael W. Mahoney. A randomized algorithm for a tensor-based generalization of the singular value decomposition. *Linear Algebra and Its Applications*, 420(2-3):553–571, 2007.
- [159] Zlatko Drmac and Serkan Gugercin. A new selection operator for the discrete empirical interpolation method – improved a priori error bound and extensions. *SIAM Journal on Scientific Computing*, 38(2):A631–A648, 2016.
- [160] Qiang Du and Max Gunzburger. Model reduction by proper orthogonal decomposition coupled with centroidal voronoi tessellations (keynote). In *ASME 2002 Joint US-European Fluids Engineering Division Conference*, pages 1401–1406. American Society of Mechanical Engineers, 2002.
- [161] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, 2000. [162] Jed A. Duersch and Ming Gu. Randomized QR with column pivoting. *SIAM Journal on Scientific Computing*, 39(4):C263–C291, 2017.
- [163] D. Duke, D. Honnery, and J. Soria. Experimental investigation of nonlinear instabilities in annular liquid sheets. *Journal of Fluid Mechanics*, 691:594–604, 2012.
- [164] Daniel Duke, Julio Soria, and Damon Honnery. An error analysis of the dynamic mode decomposition. *Experiments in Fluids*, 52(2):529–542, 2012.
- [165] Geir E. Dullerud and Fernando Paganini. *A Course in Robust Control Theory: A Convex Approach*. Texts in Applied Mathematics. Springer, Berlin, Heidelberg, 2000.
- [166] R. Dunne and B. J. McKeon. Dynamic stall on a pitching and surging airfoil. *Experiments in Fluids*, 56(8):1–15, 2015.
- [167] T. Duriez, S. L. Brunton, and B. R. Noack. *Machine Learning Control: Taming Nonlinear Dynamics and Turbulence*. Springer, 2016.
- [168] Thomas Duriez, Vladimir Parezanovic', Laurent Cordier, Bernd R. Noack, Joël Delville, Jean-Paul Bonnet, Marc Segond, and Markus Abel. Closed-loop turbulence control using machine learning. *arXiv preprint arXiv:1404.4589*, 2014.
- [169] Thomas Duriez, Vladimir Parezanovic, Jean-Charles Laurentie, Carine Fourment, Joël Delville, Jean-Paul Bonnet, Laurent Cordier, Bernd R Noack, Marc Segond, Markus Abel, Nicolas Gautier, Jean-Luc Aider, Cedric Raibaud, Christophe Cuvier, Michel Stanislas, and Steven L Brunton. Closed-loop control of experimental shear flows using machine learning. *AIAA Paper 2014-2219*, 7th Flow Control Conference, 2014.
- [170] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.
- [171] Jens L. Eftang, Anthony T. Patera, and Einar M. Rønquist. An “hp” certified reduced basis method for parametrized elliptic partial differential equations. *SIAM Journal on Scientific Computing*, 32(6):3170–3200, 2010.
- [172] Jeffrey L. Elman. Finding structure in time. *Cognitive Science*, 14(2):179–211, 1990.

- [173] Utku Eren, Anna Prach, Bas, aran Bahadır Koçer, Saša V. Raković, Erdal Kayacan, and Behçet Açıkmeşe. Model predictive control in aerospace systems: Current state and opportunities. *Journal of Guidance, Control, and Dynamics*, 40(7):1541–1566, 2017.
- [174] N. B. Erichson, S. L. Brunton, and J. N. Kutz. Compressed dynamic mode decomposition for real-time object detection. *Journal of Real-Time Image Processing*, 2016.
- [175] N. B. Erichson, S. L. Brunton, and J. N. Kutz. Randomized dynamic mode decomposition. *arXiv preprint arXiv:1702.02912*, 2017.
- [176] N. B. Erichson, K. Manohar, S. L. Brunton, and J. N. Kutz. Randomized CP tensor decomposition. *arXiv preprint arXiv:1703.09074*.
- [177] N. B. Erichson, S. Voronin, S. L. Brunton, and J. N. Kutz. Randomized matrix decompositions using R. *arXiv preprint arXiv:1608.02148*, 2016.
- [178] Trishan Eram, Jonathan W. Kimball, Philip T Krein, Patrick L. Chapman, and Pallab Midya. Dynamic maximum power point tracking of photovoltaic arrays using ripple correlation control. *Ieee Transactions On Power Electronics*, 21(5):1282–1291, September 2006.
- [179] Richard Everson and Lawrence Sirovich. Karhunen–Loeve procedure for gappy data. *JOSA A*, 12(8):1657–1664, 1995.
- [180] N. Fabbiane, O. Semeraro, S. Bagheri, and D. S. Henningson. Adaptive and model-based control theory applied to convectively unstable flows. *Appl. Mech. Rev.*, 66(6):060801–1–20, 2014.
- [181] B. F. Feeny. On proper orthogonal co-ordinates as indicators of modal activity. *Journal of Sound and Vibration*, 255(5):805–817, 2002.
- [182] Ronald A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Human Genetics*, 7(2):179–188, 1936.
- [183] Ronald Aylmer Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222:309–368, 1922.
- [184] P. J. Fleming and R. C. Purshouse. Evolutionary algorithms in control systems engineering: a survey. *Control Engineering Practice*, 10:1223–1241, 2002.
- [185] Jean Baptiste Joseph Fourier. *The Analytical Theory of Heat*. The University Press, 1878.
- [186] Joseph Fourier. *Theorie analytique de la chaleur, par M. Fourier*. Chez Firmin Didot, père et fils, 1822.
- [187] J. E. Fowler. Compressive-projection principal component analysis. *IEEE Transactions on Image Processing*, 18(10):2230–2242, 2009.
- [188] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.
- [189] Jerome H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, pages 1189–1232, 2001.
- [190] Alan Frieze, Ravi Kannan, and Santosh Vempala. Fast Monte-Carlo algorithms for finding low-rank approximations. *Journal of the ACM*, 51(6):1025–1041, 2004.

- [191] Xing Fu, Steven L. Brunton, and J. Nathan Kutz. Classification of birefringence in mode-locked fiber lasers using machine learning and sparse representation. *Optics Express*, 22(7):8585–8597, 2014.
- [192] Koji Fukagata, Stefan Kern, Philippe Chatelain, Petros Koumoutsakos, and Nobuhide Kasagi. Evolutionary optimization of an anisotropic compliant surface for turbulent friction drag reduction. *Journal of Turbulence*, 9(35):1–17, 2008.
- [193] F. Fukushima. A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetic*, 36:193–202, 1980.
- [194] H. Gao, J. Lam, C. Wang, and Y. Wang. Delay-dependent output-feedback stabilisation of discrete-time systems with time-varying state delay. *IEE Proceedings-Control Theory and Applications*, 151(6):691–698, 2004.
- [195] Carlos E. Garcia, David M. Prett, and Manfred Morari. Model predictive control: theory and practice: A survey. *Automatica*, 25(3):335–348, 1989.
- [196] Jorge L. Garriga and Masoud Soroush. Model predictive control tuning methods: A review. *Industrial & Engineering Chemistry Research*, 49(8):3505–3515, 2010.
- [197] Carl-Friedrich Gauss. *Theoria combinationis observationum erroribus minimis obnoxiae*, volume 1. Henricus Dieterich, 1823.
- [198] C. F. Gauss. Nachlass: Theoria interpolationis methodo nova tractata, volume werke. *Königliche Gesellschaft der Wissenschaften, Göttingen*, 1866.
- [199] Nicolas Gautier, J-L. Aider, Thomas Duriez, B. R. Noack, Marc Segond, and Markus Abel. Closed-loop separation control using machine learning. *Journal of Fluid Mechanics*, 770:442–457, 2015.
- [200] M. Gavish and D. L. Donoho. The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Transactions on Information Theory*, 60(8):5040–5053, 2014.
- [201] Mattia Gazzola, Oleg V. Vasilyev, and Petros Koumoutsakos. Shape optimization for drag reduction in linked bodies using evolution strategies. *Computers & Structures*, 89(11):1224–1231, 2011.
- [202] Gregor Gelbert, Jonas P. Moeck, Christian O. Paschereit, and Rudibert King. Advanced algorithms for gradient estimation in one- and two-parameter extremum seeking controllers. *Journal of Process Control*, 22(4):700–709, 2012.
- [203] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 23(6):643–660, 2001.
- [204] Jan J. Gerbrands. On the relationships between SVD, KLT and PCA. *Pattern Recognition*, 14(1):375–381, 1981.
- [205] A. C. Gilbert and P. Indyk. Sparse recovery using sparse matrices. *Proceedings of the IEEE*, 98(6):937–947, 2010.
- [206] A. C. Gilbert, J. Y. Park, and M. B. Wakin. Sketched SVD: Recovering spectral features from compressive measurements. *ArXiv e-prints*, 2012.
- [207] A. C. Gilbert, M. J. Strauss, and J. A. Tropp. A tutorial on fast Fourier sampling. *IEEE Signal Processing Magazine*, pages 57–66, 2008.
- [208] Bryan Glaz, Li Liu, and Peretz P. Friedmann. Reduced-order nonlinear unsteady aerodynamic modeling using a surrogate-based recurrence framework. *AIAA Journal*, 48(10):2418–2429, 2010.

- [209] Philip J Goddard and Keith Glover. Controller approximation: approaches for preserving H^∞ performance. *IEEE Transactions on Automatic Control*, 43(7):858–871, 1998.
- [210] David E Goldberg. *Genetic Algorithms*. Pearson Education India, 2006.
- [211] G. H. Golub and C. Reinsch. Singular value decomposition and least squares solutions. *Numerical Mathematics*, 14:403–420, 1970.
- [212] Gene Golub and William Kahan. Calculating the singular values and pseudo-inverse of a matrix. *Journal of the Society for Industrial & Applied Mathematics, Series B: Numerical Analysis*, 2(2):205–224, 1965.
- [213] Gene Golub, Stephen Nash, and Charles Van Loan. A Hessenberg-Schur method for the problem $ax + xb = c$. *IEEE Transactions on Automatic Control*, 24(6):909–913, 1979.
- [214] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*, volume 3. JHU Press, 2012.
- [215] R. Gonzalez-Garcia, R. Rico-Martinez, and I. G. Kevrekidis. Identification of distributed parameter systems: A neural net based approach. *Comp. & Chem. Eng.*, 22:S965–S968, 1998.
- [216] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [217] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.
- [218] Michael Grant, Stephen Boyd, and Yinyu Ye. *Cvx: Matlab software for disciplined convex programming*, 2008.
- [219] Alex Graves, Greg Wayne, and Ivo Danihelka. Neural Turing machines. *arXiv preprint arXiv:1410.5401*, 2014.
- [220] Anne Greenbaum. *Iterative Methods for Solving Linear Systems*. SIAM, 1997.
- [221] Mohinder S. Grewal. Kalman filtering. In *International Encyclopedia of Statistical Science*, pages 705–708. Springer, 2011.
- [222] M. Grilli, P. J. Schmid, S. Hickel, and N. A. Adams. Analysis of unsteady behaviour in shockwave turbulent boundary layer interaction. *Journal of Fluid Mechanics*, 700:16–28, 2012.
- [223] Jacob Grosek and J. Nathan Kutz. Dynamic mode decomposition for real-time background/foreground separation in video. *arXiv preprint arXiv:1404.7592*, 2014.
- [224] Ming Gu. Subspace iteration randomization and singular value problems. *SIAM Journal on Scientific Computing*, 37(3):1139–1173, 2015.
- [225] F. Gueniat, L. Mathelin, and L. Pastur. A dynamic mode decomposition approach for large and arbitrarily sampled systems. *Physics of Fluids*, 27(2):025113, 2015.
- [226] Fredrik Gustafsson, Fredrik Gunnarsson, Niclas Bergman, Urban Forssell, Jonas Jansson, Rickard Karlsson, and P-J. Nordlund. Particle filters for positioning, navigation, and tracking. *IEEE Transactions on signal processing*, 50(2):425–437, 2002.
- [227] Alfred Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69(3):331–371, 1910.

- [228] N. Halko, P. G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011.
- [229] Nathan Halko, Per-Gunnar Martinsson, Yoel Shkolnisky, and Mark Tygert. An algorithm for the principal component analysis of large data sets. *SIAM Journal on Scientific Computing*, 33:2580–2594, 2011.
- [230] Nathan Halko, Per-Gunnar Martinsson, and Joel A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011.
- [231] Sven J. Hammarling. Numerical solution of the stable, non-negative definite Lyapunov equation. *IMA Journal of Numerical Analysis*, 2(3):303–323, 1982.
- [232] Sangbo Han and Brian Feeny. Application of proper orthogonal decomposition to structural vibration analysis. *Mechanical Systems and Signal Processing*, 17(5):989–1001, 2003.
- [233] Nikolaus Hansen, André S. P. Niederberger, Lino Guzzella, and Petros Koumoutsakos. A method for handling uncertainty in evolutionary optimization with an application to feedback control of combustion. *IEEE Transactions on Evolutionary Computation*, 13(1):180–197, 2009.
- [234] David Harrison Jr. and Daniel L. Rubinfeld. Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5(1):81–102, 1978.
- [235] Richard A. Harshman. Foundations of the PARAFAC procedure: Models and conditions for an “explanatory” multi-modal factor analysis. *UCLA working papers in phonetics*, 16:1–84, 1970. Available at www.psychology.uwo.ca/faculty/harshman/wpppfac0.pdf.
- [236] Trevor Hastie, Robert Tibshirani, Jerome Friedman, T Hastie, J. Friedman, and R. Tibshirani. *The Elements of Statistical Learning*, volume 2. Springer, 2009.
- [237] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [238] M. T. Heath, A. J. Laub, C. C. Paige, and R. C. Ward. Computing the singular value decomposition of a product of two matrices. *SIAM Journal on Scientific and Statistical Computing*, 7(4):1147–1159, 1986.
- [239] Michael Heideman, Don Johnson, and C Burrus. Gauss and the history of the fast Fourier transform. *IEEE ASSP Magazine*, 1(4):14–21, 1984.
- [240] Werner Heisenberg. Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik. In *Original Scientific Papers Wissenschaftliche Originalarbeiten*, pages 478–504. Springer, 1985.
- [241] Maziar S. Hemati, Clarence W. Rowley, Eric A. Deem, and Louis N. Cattafesta. De-biasing the dynamic mode decomposition for applied Koopman spectral analysis. *Theoretical and Computational Fluid Dynamics*, 31(4):349–368, 2017.
- [242] Maziar S. Hemati, Matthew O. Williams, and Clarence W. Rowley. Dynamic mode decomposition for large and streaming datasets. *Physics of Fluids (1994-present)*, 26(11):111701, 2014.
- [243] Kyle K. Herrity, Anna C. Gilbert, and Joel A. Tropp. Sparse approximation via iterative thresholding. In *Acoustics, Speech and Signal Processing, 2006*.

- ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 3, pages III–III. IEEE, 2006.
- [244] Jan S. Hesthaven, Gianluigi Rozza, and Benjamin Stamm. Certified reduced basis methods for parametrized partial differential equations. *SpringerBriefs in Mathematics*, 2015.
- [245] Tony Hey, Stewart Tansley, Kristin M. Tolle, et al. *The Fourth Paradigm: Data-Intensive Scientific Discovery*, volume 1. Microsoft research Redmond, WA, 2009.
- [246] Geoffrey E. Hinton and Terrence J. Sejnowski. Learning and relearning in boltzmann machines. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 1(282-317):2, 1986.
- [247] B. L. Ho and R. E. Kalman. Effective construction of linear state-variable models from input/output data. In *Proceedings of the 3rd Annual Allerton Conference on Circuit and System Theory*, pages 449–459, 1965.
- [248] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.
- [249] Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.
- [250] John H. Holland. *Adaptation in natural and artificial systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. University of Michigan Press, 1975.
- [251] P. Holmes, J. L. Lumley, G. Berkooz, and C. W. Rowley. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press, Cambridge, 2nd paperback edition, 2012.
- [252] Philip Holmes and John Guckenheimer. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, volume 42 of *Applied Mathematical Sciences*. Springer-Verlag, Berlin, Heidelberg, 1983.
- [253] Eberhard Hopf. The partial differential equation $ut + uux = \mu uxx$. *Communications on Pure and Applied Mathematics*, 3(3):201–230, 1950.
- [254] John J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.
- [255] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.
- [256] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, September 1933.
- [257] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:498–520, October 1933.
- [258] C. Huang, W. E. Anderson, M. E. Harvazinski, and V. Sankaran. Analysis of self-excited combustion instabilities using decomposition techniques. In *51st AIAA Aerospace Sciences Meeting*, pages 1–18, 2013.
- [259] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154, 1962.
- [260] Peter J. Huber. Robust statistics. In *International Encyclopedia of Statistical Science*, pages 1248–1251. Springer, 2011.

- [261] S. J. Illingworth, A. S. Morgans, and C. W. Rowley. Feedback control of flow resonances using balanced reduced-order models. *Journal of Sound and Vibration*, 330(8):1567–1581, 2010.
- [262] Eric Jacobsen and Richard Lyons. The sliding DFT. *IEEE Signal Processing Magazine*, 20(2):74–80, 2003.
- [263] Herbert Jaeger and Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667):78–80, 2004.
- [264] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning*. Springer, 2013.
- [265] M. C. Johnson, S. L. Brunton, N. B. Kundtz, and J. N. Kutz. Extremum-seeking control of a beam pattern of a reconfigurable holographic metamaterial antenna. *Journal of the Optical Society of America A*, 33(1):59–68, 2016.
- [266] Richard A. Johnson and Dean Wichern. *Multivariate Analysis*. Wiley Online Library, 2002.
- [267] W. B Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. *Contemporary Mathematics*, 26(189-206):1, 1984.
- [268] Ian Jolliffe. *Principal Component Analysis*. Wiley Online Library, 2005.
- [269] Siddharth Joshi and Stephen Boyd. Sensor selection via convex optimization. *IEEE Transactions on Signal Processing*, 57(2):451–462, 2009.
- [270] Mihailo R. Jovanović, Peter J. Schmid, and Joseph W. Nichols. Sparsity-promoting dynamic mode decomposition. *Physics of Fluids*, 26(2):024103, 2014.
- [271] J. N. Juang. *Applied System Identification*. Prentice Hall PTR, Upper Saddle River, New Jersey, 1994.
- [272] J. N. Juang and R. S. Pappa. An eigensystem realization algorithm for modal parameter identification and model reduction. *Journal of Guidance, Control, and Dynamics*, 8(5):620–627, 1985.
- [273] J. N. Juang, M. Phan, L. G. Horta, and R. W. Longman. Identification of observer/Kalman filter Markov parameters: Theory and experiments. Technical Memorandum 104069, NASA, 1991.
- [274] Simon J. Julier and Jeffrey K. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *Int. symp. aerospace/defense sensing, simul. and controls*, volume 3, pages 182–193. Orlando, FL, 1997.
- [275] Simon J. Julier and Jeffrey K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, 2004.
- [276] E. Kaiser, J. N. Kutz, and S. L. Brunton. Data-driven discovery of Koopman eigenfunctions for control. *arXiv preprint arXiv:1707.01146*, 2017.
- [277] Eurika Kaiser, J. Nathan Kutz, and Steven L. Brunton. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. To appear in *Proceedings of the Royal Society A*. *arXiv preprint arXiv:1711.05501*, 2017.
- [278] Eurika Kaiser, Bernd R. Noack, Laurent Cordier, Andreas Spohn, Marc Segond, Markus Abel, Guillaume Daviller, Jan Östh, Siniša Krajnović, and Robert K Niven. Cluster-based reduced-order modelling of a mixing layer. *Journal of Fluid Mechanics*, 754:365–414, 2014.
- [279] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1):35–45, 1960.

- [280] K. Karhunen. Über lineare methoden in der wahrscheinlichkeitsrechnung, vol. 37. *Annales AcademiæScientiarum Fennicæ, Ser. A. I*, 1947.
- [281] Kévin Kasper, Lionel Mathelin, and Hisham Abou-Kandil. A machine learning approach for constrained sensor placement. In *American Control Conference (ACC), 2015*, pages 4479–4484. IEEE, 2015.
- [282] A. K. Kassam and L. N. Trefethen. Fourth-order time-stepping for stiff PDEs. *SIAM Journal on Scientific Computing*, 26(4):1214–1233, 2005.
- [283] Michael Kearns and Leslie Valiant. Cryptographic limitations on learning boolean formulae and finite automata. *Journal of the ACM (JACM)*, 41(1):67–95, 1994.
- [284] Anthony R. Kellems, Saifon Chaturantabut, Danny C. Sorensen, and Steven J. Cox. Morphologically accurate reduced order modeling of spiking neurons. *Journal of Computational Neuroscience*, 28(3):477–494, 2010.
- [285] J. Kepler. *Tabulae Rudolphinae, quibus Astronomicae scientiae, temporum longinquitate collapsae Restauratio continetur*. Ulm: Jonas Saur, 1627.
- [286] Gaëtan Kerschen and Jean-Claude Golinval. Physical interpretation of the proper orthogonal modes using the singular value decomposition. *Journal of Sound and Vibration*, 249(5):849–865, 2002.
- [287] Gaëtan Kerschen, Jean-claude Golinval, Alexander F. Vakakis, and Lawrence A. Bergman. The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: an overview. *Nonlinear Dynamics*, 41(1-3):147–169, 2005.
- [288] I. G. Kevrekidis, C. W. Gear, J. M. Hyman, P. G. Kevrekidis, O. Runborg, and C. Theodoropoulos. Equation-free, coarse-grained multiscale computation: Enabling microscopic simulators to perform system-level analysis. *Communications in Mathematical Science*, 1(4):715–762, 2003.
- [289] N. J. Killingsworth and M. Krstic. PID tuning using extremum seeking: online, model-free performance optimization. *IEEE Control Systems Magazine*, February:70–79, 2006.
- [290] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [291] M. Kirby and L. Sirovich. Application of the Karhunen-Loève procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 12(1):103–108, 1990.
- [292] V. C. Klema and A. J. Laub. The singular value decomposition: Its computation and some applications. *IEEE Transactions on Automatic Control*, 25(2):164–176, 1980.
- [293] Stefan Klus, Feliks Nüske, Péter Koltai, Hao Wu, Ioannis Kevrekidis, Christof Schütte, and Frank Noé. Data-driven model reduction and transfer operator approximation. *Journal of Nonlinear Science*, pages 1–26, 2018.
- [294] Richard Koch. *The 80/20 Principle*. Nicholas Brealey Publishing, 1997.
- [295] Richard Koch. *Living the 80/20 Way*. Audio-Tech Business Book Summaries, Incorporated, 2006.
- [296] Richard Koch. *The 80/20 Principle: The Secret to Achieving More with Less*. Crown Business, 2011.
- [297] Richard Koch. *The 80/20 Principle and 92 other Powerful Laws of Nature: the Science of Success*. Nicholas Brealey Publishing, 2013.

- [298] Teuvo Kohonen. The self-organizing map. *Neurocomputing*, 21(1-3):1–6, 1998.
- [299] Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, September 2009.
- [300] B. O. Koopman. Hamiltonian systems and transformation in Hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, 1931.
- [301] B. O. Koopman and J.-v. Neumann. Dynamical systems of continuous spectra. *Proceedings of the National Academy of Sciences of the United States of America*, 18(3):255, 1932.
- [302] Milan Korda and Igor Mezic'. Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control. *Automatica*, 93:149–160, 2018.
- [303] Milan Korda and Igor Mezic'. On convergence of extended dynamic mode decomposition to the Koopman operator. *Journal of Nonlinear Science*, 28(2):687–710, 2018.
- [304] Petros Koumoutsakos, Jonathan Freund, and David Parekh. Evolution strategies for automatic optimization of jet mixing. *AIAA Journal*, 39(5):967–969, 2001.
- [305] Krzysztof Kowalski, Willi-Hans Steeb, and K. Kowalksi. *Nonlinear Dynamical Systems and Carleman Linearization*. World Scientific, 1991.
- [306] John R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, volume 1. MIT press, 1992.
- [307] John R. Koza, Forrest H. Bennett III, and Oscar Stiffelman. Genetic programming as a darwinian invention machine. In *Genetic Programming*, pages 93–108. Springer, 1999.
- [308] Boris Kramer, Piyush Grover, Petros Boufounos, Mouhacine Benosman, and Saleh Nabi. Sparse sensing and dmd based identification of flow regimes and bifurcations in complex flows. *SIAM Journal on Applied Dynamical Systems*, 16(2):1164–1196, 2017.
- [309] J. P. Krieger and M. Krstic. Extremum seeking based on atmospheric turbulence for aircraft endurance. *Journal of Guidance, Control, and Dynamics*, 34(6):1876–1885, 2011.
- [310] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.
- [311] M. Krstic, A. Krupadanam, and C. Jacobson. Self-tuning control of a nonlinear model of combustion instabilities. *IEEE Tr. Contr. Syst. Technol.*, 7(4):424–436, 1999.
- [312] M. Krstic' and H. H. Wang. Stability of extremum seeking feedback for general nonlinear dynamic systems. *Automatica*, 36:595–601, 2000.
- [313] Tejas D Kulkarni, William F Whitney, Pushmeet Kohli, and Josh Tenenbaum. Deep convolutional inverse graphics network. In *Advances in Neural Information Processing Systems*, pages 2539–2547, 2015.
- [314] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.
- [315] Karl Kunisch and Stefan Volkwein. Optimal snapshot location for computing pod basis functions. *ESAIM: Mathematical Modelling and Numerical Analysis*, 44(3):509–529, 2010.

- [316] J. N. Kutz. *Data-Driven Modeling & Scientific Computation: Methods for Complex Systems & Big Data*. Oxford University Press, 2013.
- [317] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. SIAM, 2016.
- [318] J. N. Kutz, X. Fu, and S. L. Brunton. Multi-resolution dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 15(2):713–735, 2016.
- [319] J. Nathan Kutz, Syuzanna Sargsyan, and Steven L Brunton. Leveraging sparsity and compressive sensing for reduced order modeling. In *Model Reduction of Parametrized Systems*, pages 301–315. Springer, 2017.
- [320] Sanjay Lall, Jerrold E. Marsden, and Sonja Glavaški. Empirical model reduction of controlled nonlinear systems. In *IFAC World Congress*, volume F, pages 473–478. International Federation of Automatic Control, 1999.
- [321] Sanjay Lall, Jerrold E. Marsden, and Sonja Glavaški. A subspace approach to balanced truncation for model reduction of nonlinear control systems. *International Journal of Robust and Nonlinear Control*, 12(6):519–535, 2002.
- [322] Yueheng Lan and Igor Mezic. Linearization in the large of nonlinear systems and Koopman operator spectrum. *Physica D: Nonlinear Phenomena*, 242(1):42–53, 2013.
- [323] Alan Laub. A Schur method for solving algebraic Riccati equations. *IEEE Transactions on automatic control*, 24(6):913–921, 1979.
- [324] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436, 2015.
- [325] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [326] Jay H. Lee. Model predictive control: Review of the three decades of development. *International Journal of Control, Automation and Systems*, 9(3):415–424, 2011.
- [327] K.C. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 27(5):684–698, 2005.
- [328] Adrien Marie Legendre. *Nouvelles méthodes pour la détermination des orbites des comètes*. F. Didot, 1805.
- [329] V. Lenaerts, Gaëtan Kerschen, and Jean-Claude Golinval. Proper orthogonal decomposition for model updating of nonlinear mechanical systems. *Mechanical Systems and Signal Processing*, 15(1):31–43, 2001.
- [330] Ian Lenz, Ross A. Knepper, and Ashutosh Saxena. Deepmpc: Learning deep latent features for model predictive control. In *Robotics: Science and Systems*, 2015.
- [331] R. Leyva, C. Alonso, I. Queinnec, A. Cid-Pastor, D. Lagrange, and L. Martinez-Salamero. MPPT of photovoltaic systems using extremum-seeking control. *Ieee Transactions On Aerospace and Electronic Systems*, 42(1):249–258, January 2006.
- [332] Qianxiao Li, Felix Dietrich, Erik M. Bollt, and Ioannis G. Kevrekidis. Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the Koopman operator. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(10):103111, 2017.

- [333] Y. C. Liang, H. P. Lee, S. P. Lim, W. Z. Lin, K. H. Lee, and C. G. Wu. Proper orthogonal decomposition and its applications-part i: Theory. *Journal of Sound and vibration*, 252(3):527–544, 2002.
- [334] Edo Liberty. Simple and deterministic matrix sketching. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 581–588. ACM, 2013.
- [335] Edo Liberty, Franco Woolfe, Per-Gunnar Martinsson, Vladimir Rokhlin, and Mark Tygert. Randomized algorithms for the low-rank approximation of matrices. *Proceedings of the National Academy of Sciences*, 104:20167–20172, 2007.
- [336] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.
- [337] Zhouchen Lin, Minming Chen, and Yi Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010.
- [338] L. Ljung. *System Identification: Theory for the User*. Prentice Hall, 1999.
- [339] Stuart Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.
- [340] M. Loeve. *Probability Theory*. Van Nostrand, Princeton, NJ, 1955.
- [341] J.-C. Loiseau and S. L. Brunton. Constrained sparse Galerkin regression. *Journal of Fluid Mechanics*, 838:42–67, 2018.
- [342] J.-C. Loiseau, B. R. Noack, and S. L. Brunton. Sparse reduced-order modeling: sensor-based dynamics to full-state estimation. *Journal of Fluid Mechanics*, 844:459–490, 2018.
- [343] Richard W. Longman. Iterative learning control and repetitive control for engineering practice. *International Journal of Control*, 73(10):930–954, 2000.
- [344] E. N. Lorenz. Empirical orthogonal functions and statistical weather prediction. Technical report, Massachusetts Institute of Technology, December 1956.
- [345] Edward N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, 1963.
- [346] D. M. Luchtenburg and C. W. Rowley. Model reduction using snapshot-based realizations. *Bulletin of the American Physical Society*, 56, 2011.
- [347] J. L. Lumley. Toward a turbulent constitutive relation. *Journal of Fluid Mechanics*, 41(02):413–434, 1970.
- [348] Bethany Lusch, Eric C. Chi, and J. Nathan Kutz. Shape constrained tensor decompositions using sparse representations in over-complete libraries. *arXiv preprint arXiv:1608.04674*, 2016.
- [349] Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*. *arXiv preprint arXiv:1712.09707*, 2018.
- [350] F. Lusseyran, F. Gueniat, J. Basley, C. L. Douay, L. R. Pastur, T. M. Faure, and P. J. Schmid. Flow coherent structures and frequency signature: application of the dynamic modes decomposition to open cavity flow. In *Journal of Physics: Conference Series*, volume 318, page 042036. IOP Publishing, 2011.
- [351] Z. Ma, S. Ahuja, and C. W. Rowley. Reduced order models for control of fluids using the eigensystem realization algorithm. *Theor. Comput. Fluid Dyn.*, 25(1):233–247, 2011.

- [352] Wolfgang Maass, Thomas Natschläger, and Henry Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, 14(11):2531–2560, 2002.
- [353] Alan Mackey, Hayden Schaeffer, and Stanley Osher. On the compressive spectral method. *Multiscale Modeling & Simulation*, 12(4):1800–1827, 2014.
- [354] Michael W. Mahoney. Randomized algorithms for matrices and data. *Foundations and Trends in Machine Learning*, 3:123–224, 2011.
- [355] Andrew J. Majda and John Harlim. Physics constrained nonlinear regression models for time series. *Nonlinearity*, 26(1):201, 2012.
- [356] Andrew J. Majda and Yoonsang Lee. Conceptual dynamical models for turbulence. *Proceedings of the National Academy of Sciences*, 111(18):6548–6553, 2014.
- [357] Stéphane Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1999.
- [358] Stéphane Mallat. Understanding deep convolutional networks. *Phil. Trans. R. Soc. A*, 374(2065):20150203, 2016.
- [359] Stéphane G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.
- [360] John Mandel. Use of the singular value decomposition in regression analysis. *The American Statistician*, 36(1):15–24, 1982.
- [361] Niall M. Mangan, Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Inferring biological networks by sparse identification of nonlinear dynamics. *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, 2(1):52–63, 2016.
- [362] Niall M. Mangan, J. Nathan Kutz, Steven L. Brunton, and Joshua L. Proctor. Model selection for dynamical systems via sparse regression and information criteria. *Proceedings of the Royal Society A*, 473(2204):1–16, 2017.
- [363] Jordan Mann and J. Nathan Kutz. Dynamic mode decomposition for financial trading strategies. *Quantitative Finance*, pages 1–13, 2016.
- [364] K. Manohar, S. L. Brunton, and J. N. Kutz. Environmental identification in flight using sparse approximation of wing strain. *Journal of Fluids and Structures*, 70:162–180, 2017.
- [365] K. Manohar, J. N. Kutz, and S. L. Brunton. Greedy Sensor and Actuator Placement Using Balanced Model Reduction. *Bulletin of the American Physical Society*, 2018.
- [366] Krithika Manohar, Bingni W. Brunton, J. Nathan Kutz, and Steven L. Brunton. Data-driven sparse sensor placement. *IEEE Control Systems Magazine*, 38:63–86, 2018.
- [367] Krithika Manohar, Eurika Kaiser, S. L. Brunton, and J. N. Kutz. Optimized sampling for multi-scale dynamics. *SIAM Multiscale Modeling and Simulation*. *arXiv preprint arXiv:1712.05085*, 2017.
- [368] Andreas Mardt, Luca Pasquali, Hao Wu, and Frank Noé. VAMPnets: Deep learning of molecular kinetics. *Nature Communications*, 9(1), 2018.
- [369] J. E. Marsden and T. S. Ratiu. *Introduction to Mechanics and Symmetry*. Springer-Verlag, 2nd edition, 1999.
- [370] Per-Gunnar Martinsson. Randomized methods for matrix computations and analysis of high dimensional data. *arXiv preprint arXiv:1607.01649*, 2016.

- [371] Per-Gunnar Martinsson, Vladimir Rokhlin, and Mark Tygert. A randomized algorithm for the decomposition of matrices. *Applied and Computational Harmonic Analysis*, 30:47–68, 2011.
- [372] John L. Maryak, James C. Spall, and Bryan D. Heydon. Use of the Kalman filter for inference in state-space models with unknown noise distributions. *IEEE Transactions on Automatic Control*, 49(1):87–90, 2004.
- [373] L. Massa, R. Kumar, and P. Ravindran. Dynamic mode decomposition analysis of detonation waves. *Physics of Fluids (1994-present)*, 24(6):066101, 2012.
- [374] Lionel Mathelin, Kévin Kasper, and Hisham Abou-Kandil. Observable dictionary learning for high-dimensional statistical inference. *Archives of Computational Methods in Engineering*, 25(1):103–120, 2018.
- [375] R. Maury, M. Keonig, L. Cattafesta, P. Jordan, and J. Delville. Extremum-seeking control of jet noise. *Aeroacoustics*, 11(3&4):459–474, 2012.
- [376] I. Mezic'. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41(1-3):309–325, 2005.
- [377] I. Mezic'. Analysis of fluid flows via spectral properties of the Koopman operator. *Ann. Rev. Fluid Mech.*, 45:357–378, 2013.
- [378] I. Mezic'. *Spectral Operator Methods in Dynamical Systems: Theory and Applications*. Springer, 2017.
- [379] Igor Mezic' and Andrzej Banaszuk. Comparison of systems with complex behavior. *Physica D: Nonlinear Phenomena*, 197(1):101–133, 2004.
- [380] Igor Mezic' and Stephen Wiggins. A method for visualization of invariant sets of dynamical systems based on the ergodic partition. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 9(1):213–218, 1999.
- [381] Michele Milano and Petros Koumoutsakos. Neural network modeling for near wall turbulent flow. *Journal of Computational Physics*, 182(1):1–26, 2002.
- [382] T. M. Mitchell. *Machine Learning*. McGraw Hill, 1997.
- [383] Y. Mizuno, D. Duke, C. Atkinson, and J. Soria. Investigation of wall-bounded turbulent flow using dynamic mode decomposition. In *Journal of Physics: Conference Series*, volume 318, page 042040. IOP Publishing, 2011.
- [384] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *International Conference on Machine Learning*, pages 1928–1937, 2016.
- [385] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.
- [386] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529, 2015.
- [387] J. P. Moeck, J.-F. Bourgouin, D. Durox, T. Schuller, and S. Candel. Tomographic reconstruction of heat release rate perturbations induced by helical modes in turbulent swirl flames. *Experiments in Fluids*, 54(4):1–17, 2013.
- [388] B. C. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control*, AC-26(1):17–32, 1981.

- [389] Calvin C. Moore. Ergodic theorem, ergodic theory, and statistical mechanics. *Proceedings of the National Academy of Sciences*, 112(7):1907–1911, 2015.
- [390] Kevin L. Moore. *Iterative Learning Control for Deterministic Systems*. Springer Science & Business Media, 2012.
- [391] Manfred Morari and Jay H. Lee. Model predictive control: past, present and future. *Computers & Chemical Engineering*, 23(4):667–682, 1999.
- [392] T. W. Muld, G. Efraimsson, and D. S. Henningson. Flow structures around a high-speed train extracted using proper orthogonal decomposition and dynamic mode decomposition. *Computers & Fluids*, 57:87–97, 2012.
- [393] T. W. Muld, G. Efraimsson, and D. S. Henningson. Mode decomposition on surface-mounted cube. *Flow, Turbulence and Combustion*, 88(3):279–310, 2012.
- [394] S. D. Müller, M Milano, and P. Koumoutsakos. Application of machine learning algorithms to flow modeling and optimization. *Annual Research Briefs*, pages 169–178, 1999.
- [395] Iulian Munteanu, Antoneta Iuliana Bratcu, and Emil Ceanga. Wind turbulence used as searching signal for MPPT in variable-speed wind energy conversion systems. *Renewable Energy*, 34(1):322–327, January 2009.
- [396] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT press, 2012.
- [397] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.
- [398] D. Needell and J. A. Tropp. CoSaMP: iterative signal recovery from incomplete and inaccurate samples. *Communications of the ACM*, 53(12):93–100, 2010.
- [399] J. v Neumann. Proof of the quasi-ergodic hypothesis. *Proceedings of the National Academy of Sciences*, 18(1):70–82, 1932.
- [400] N. C. Nguyen, A. T. Patera, and J. Peraire. A best points interpolation method for efficient approximation of parametrized functions. *International Journal for Numerical Methods in Engineering*, 73(4):521–543, 2008.
- [401] Yves Nievergelt and Y. Nievergelt. *Wavelets Made Easy*, volume 174. Springer, 1999.
- [402] B. R. Noack, K. Afanasiev, M. Morzynski, G. Tadmor, and F. Thiele. A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *Journal of Fluid Mechanics*, 497:335–363, 2003.
- [403] B. R. Noack, T. Duriez, L. Cordier, M. Segond, M. Abel, S. L. Brunton, M. Morzynski, J.-C. Laurentie, V. Parezanovic, and J.-P. Bonnet. Closed-loop turbulence control with machine learning methods. *Bulletin Am. Phys. Soc.*, 58(18):M25.0009, p. 418, 2013.
- [404] Bernd R. Noack, Marek Morzynski, and Gilead Tadmor. *Reduced-Order Modeling for Flow Control*, volume 528. Springer Science & Business Media, 2011.
- [405] Frank Noé and Feliks Nuske. A variational approach to modeling slow processes in stochastic dynamical systems. *Multiscale Modeling & Simulation*, 11(2):635–655, 2013.
- [406] E. Noether. Invariante variationsprobleme nachr. d. könig. gesellsch. d. wiss. zu göttingen, math-phys. klasse 1918: 235–257. *English Reprint: physics/0503066*, <http://dx.doi.org/10.1080/00411457108231446>, page 57, 1918.

- [407] Feliks Nüske, Bettina G. Keller, Guillermo Pérez-Hernández, Antonia S. J. S. Mey, and Frank Noé. Variational approach to molecular kinetics. *Journal of Chemical Theory and Computation*, 10(4):1739–1752, 2014.
- [408] Feliks Nüske, Reinhold Schneider, Francesca Vitalini, and Frank Noé. Variational tensor approach for approximating the rare-event kinetics of macromolecular systems. *J. Chem. Phys.*, 144(5):054105, 2016.
- [409] H. Nyquist. Certain topics in telegraph transmission theory. *Transactions of the A. I. E. E.*, pages 617–644, FEB 1928.
- [410] Goro Obinata and Brian D. O. Anderson. *Model reduction for control system design*. Springer Science & Business Media, 2012.
- [411] C. M. Ostoich, D. J. Bodony, and P. H. Geubelle. Interaction of a Mach 2.25 turbulent boundary layer with a fluttering panel using direct numerical simulation. *Physics of Fluids (1994-present)*, 25(11):110806, 2013.
- [412] Samuel E. Otto and Clarence W. Rowley. Linearly-recurrent autoencoder networks for learning dynamics. *arXiv preprint arXiv:1712.01378*, 2017.
- [413] Y. Ou, C. Xu, E. Schuster, T. C. Luce, J. R. Ferron, M. L. Walker, and D. A. Humphreys. Design and simulation of extremum-seeking open-loop optimal control of current profile in the DIII-D tokamak. *Plasma Physics and Controlled Fusion*, 50:115001–1–115001–24, 2008.
- [414] Vidvuds Ozolin, š, Rongjie Lai, Russel Caflisch, and Stanley Osher. Compressed modes for variational problems in mathematics and physics. *Proceedings of the National Academy of Sciences*, 110(46):18368–18373, 2013.
- [415] C. Pan, D. Yu, and J. Wang. Dynamical mode decomposition of Gurney flap wake flow. *Theoretical and Applied Mechanics Letters*, 1(1):012002, 2011.
- [416] V. Parezanovic, J.-C. Laurentie, T. Duriez, C. Fourment, J. Delville, J.-P. Bonnet, L. Cordier, B. R. Noack, M. Segond, M. Abel, T. Shaqarin, and S. L. Brunton. Mixing layer manipulation experiment – from periodic forcing to machine learning closed-loop control. *Journal Flow Turbulence and Combustion*, 94(1):155–173, 2015.
- [417] Vladimir Parezanovic, Thomas Duriez, Laurent Cordier, Bernd R. Noack, Joël Delville, Jean- Paul Bonnet, Marc Segond, Markus Abel, and Steven L. Brunton. Closed-loop control of an experimental mixing layer using machine learning control. *arXiv preprint arXiv:1408.3259*, 2014.
- [418] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(7–12):559–572, 1901.
- [419] B. Peherstorfer, D. Butnaru, K. Willcox, and H.-J. Bungartz. Localized discrete empirical interpolation method. *SIAM Journal on Scientific Computing*, 36(1):A168–A192, 2014.
- [420] Benjamin Peherstorfer and Karen Willcox. Detecting and adapting to parameter changes for reduced models of dynamic data-driven application systems. *Procedia Computer Science*, 51:2553–2562, 2015.
- [421] Benjamin Peherstorfer and Karen Willcox. Dynamic data-driven reduced-order models. *Computer Methods in Applied Mechanics and Engineering*, 291:21–41, 2015.
- [422] Benjamin Peherstorfer and Karen Willcox. Online adaptive model reduction for nonlinear systems via low-rank updates. *SIAM Journal on Scientific Computing*, 37(4):A2123–A2150, 2015.

- [423] Sebastian Peitz and Stefan Klus. Koopman operator-based model reduction for switched- system control of PDEs. *arXiv preprint arXiv:1710.06759*, 2017.
- [424] S. D. Pendergrass, J. N. Kutz, and S. L. Brunton. Streaming GPU singular value and dynamic mode decompositions. *arXiv preprint arXiv:1612.07875*, 2016.
- [425] Roger Penrose. A generalized inverse for matrices. In *Mathematical proceedings of the Cambridge philosophical society*, volume 51, pages 406–413. Cambridge University Press, 1955.
- [426] Roger Penrose and John Arthur Todd. On best approximate solutions of linear matrix equations. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 52, pages 17–19. Cambridge Univ Press, 1956.
- [427] Lawrence Perko. *Differential Equations and Dynamical Systems*, volume 7. Springer Science & Business Media, 2013.
- [428] M. Phan, L. G. Horta, J. N. Juang, and R. W. Longman. Linear system identification via an asymptotically stable observer. *Journal of Optimization Theory and Applications*, 79:59–86, 1993.
- [429] Mark A. Pinsky. *Introduction to Fourier analysis and wavelets*, volume 102. American Mathematical Soc., 2002.
- [430] T. Poggio. Deep learning: mathematics and neuroscience. *Views & Reviews, McGovern Center for Brains, Minds and Machines*, pages 1–7, 2016.
- [431] Philippe Poncet, Georges-Henri Cottet, and Petros Koumoutsakos. Control of three-dimensional wakes using evolution strategies. *Comptes Rendus Mecanique*, 333(1):65–77, 2005.
- [432] Christopher Poultney, Sumit Chopra, Yann L Cun, et al. Efficient learning of sparse representations with an energy-based model. In *Advances in Neural Information Processing systems*, pages 1137–1144, 2007.
- [433] J. L. Proctor, S. L. Brunton, B. W. Brunton, and J. N. Kutz. Exploiting sparsity and equation-free architectures in complex systems (invited review). *The European Physical Journal Special Topics*, 223(13):2665–2684, 2014.
- [434] Joshua L. Proctor, Steven L. Brunton, and J. Nathan Kutz. Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems*, 15(1):142–161, 2016.
- [435] Joshua L. Proctor and Philip A. Eckhoff. Discovering dynamic patterns from infectious disease data using dynamic mode decomposition. *International Health*, 7(2):139–145, 2015.
- [436] H. Qi and S. M. Hughes. Invariance of principal components under low-dimensional random projection of the data. IEEE International Conference on Image Processing, October 2012.
- [437] Shie Qian and Dapang Chen. Discrete Gabor transform. *IEEE Transactions on Signal Processing*, 41(7):2429–2438, 1993.
- [438] S. J. Qin and T. A. Badgwell. An overview of industrial model predictive control technology. In *AIChE Symposium Series*, volume 93, pages 232–256, 1997.
- [439] S. Joe Qin and Thomas A. Badgwell. A survey of industrial model predictive control technology. *Control Engineering Practice*, 11(7):733–764, 2003.
- [440] Qing Qu, Ju Sun, and John Wright. Finding a sparse vector in a subspace: Linear sparsity using alternating directions. In *Advances in Neural Information Processing Systems 27*, pages 3401–3409, 2014.

- [441] A. Quarteroni and G. Rozza. *Reduced Order Methods for Modeling and Computational Reduction*, volume 9 of *MS&A – Modeling, Simulation & Applications*. Springer, 2013.
- [442] Alfio Quarteroni, Andrea Manzoni, and Federico Negri. *Reduced Basis Methods for Partial Differential Equations: An Introduction*, volume 92. Springer, 2015.
- [443] J. Ross Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, 1986.
- [444] J Ross Quinlan. *C4. 5: Programs for Machine Learning*. Elsevier, 2014.
- [445] Maziar Raissi and George Em Karniadakis. Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, 357:125–141, 2018.
- [446] C. Radhakrishna Rao. The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2):159–203, 1948.
- [447] James B. Rawlings. Tutorial overview of model predictive control. *IEEE Control Systems*, 20(3):38–52, 2000.
- [448] Soumya Raychaudhuri, Joshua M. Stuart, and Russ B. Altman. Principal components analysis to summarize microarray experiments: application to sporulation time series. In *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, page 455. NIH Public Access, 2000.
- [449] Rolf H. Reichle, Dennis B. McLaughlin, and Dara Entekhabi. Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, 130(1):103–114, 2002.
- [450] B. Ren, P. Frihauf, R. J. Rafac, and M. Krstić. Laser pulse shaping via extremum seeking. *Control Engineering Practice*, 20:674–683, 2012.
- [451] Branko Ristic, Sanjeev Arulampalam, and Neil James Gordon. *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech house, 2004.
- [452] Anthony John Roberts. *Model Emergent Dynamics in Complex Systems*. SIAM, 2014.
- [453] Charles A. Rohde. Generalized inverses of partitioned matrices. *Journal of the Society for Industrial & Applied Mathematics*, 13(4):1033–1035, 1965.
- [454] Vladimir Rokhlin, Arthur Szlam, and Mark Tygert. A randomized algorithm for principal component analysis. *SIAM Journal on Matrix Analysis and Applications*, 31:1100–1124, 2009.
- [455] C. W. Rowley, T. Colonius, and R. M. Murray. Model reduction for compressible flows using POD and Galerkin projection. *Physica D*, 189:115–129, 2004.
- [456] C. W. Rowley, I. Mezic, S. Bagheri, P. Schlatter, and D. S. Henningson. Spectral analysis of nonlinear flows. *J. Fluid Mech.*, 645:115–127, 2009.
- [457] Clarence W. Rowley and Jerrold E. Marsden. Reconstruction equations and the Karhunen–Loève expansion for systems with symmetry. *Physica D: Nonlinear Phenomena*, 142(1):1–19, 2000.
- [458] C.W. Rowley. Model reduction for fluids using balanced proper orthogonal decomposition. *Int. J. Bifurcation and Chaos*, 15(3):997–1013, 2005.
- [459] S. Roy, J.-C. Hua, W. Barnhill, G. H. Gunaratne, and J. R. Gord. Deconvolution of reacting-flow dynamics using proper orthogonal and dynamic mode decompositions. *Physical Review E*, 91(1):013001, 2015.

- [460] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3(e1602614), 2017.
- [461] Themistoklis P. Sapsis and Andrew J. Majda. Statistically accurate low-order models for uncertainty quantification in turbulent dynamical systems. *Proceedings of the National Academy of Sciences*, 110(34):13705–13710, 2013.
- [462] S. Sargsyan, S. L. Brunton, and J. N. Kutz. Nonlinear model reduction for dynamical systems using sparse sensor locations from learned libraries. *Physical Review E*, 92(033304), 2015.
- [463] S. Sarkar, S. Ganguly, A. Dalal, P. Saha, and S. Chakraborty. Mixed convective flow stability of nanofluids past a square cylinder by dynamic mode decomposition. *International Journal of Heat and Fluid Flow*, 44:624–634, 2013.
- [464] Tamas Sarlos. Improved approximation algorithms for large matrices via random projections. In *Foundations of Computer Science. 47th Annual IEEE Symposium on*, pages 143–152, 2006.
- [465] T. Sayadi, P. J. Schmid, J. W. Nichols, and P. Moin. Reduced-order representation of near-wall structures in the late transitional boundary layer. *Journal of Fluid Mechanics*, 748:278–301, 2014.
- [466] Taraneh Sayadi and Peter J. Schmid. Parallel data-driven decomposition algorithm for large-scale datasets: with application to transitional boundary layers. *Theoretical and Computational Fluid Dynamics*, pages 1–14, 2016.
- [467] H. Schaeffer, R. Caflisch, C. D. Hauck, and S. Osher. Sparse dynamics for partial differential equations. *Proceedings of the National Academy of Sciences USA*, 110(17):6634–6639, 2013.
- [468] Hayden Schaeffer. Learning partial differential equations via data discovery and sparse optimization. In *Proc. R. Soc. A*, volume 473, page 20160446. The Royal Society, 2017.
- [469] Hayden Schaeffer and Scott G. McCalla. Sparse model selection via integral terms. *Physical Review E*, 96(2):023302, 2017.
- [470] Robert E. Schapire. The strength of weak learnability. *Machine learning*, 5(2):197–227, 1990.
- [471] M. Schlegel, B. R. Noack, and G. Tadmor. Low-dimensional Galerkin models and control of transitional channel flow. Technical Report 01/2004, Hermann-Föttinger-Institut für Strömungsmechanik, Technische Universität Berlin, Germany, 2004.
- [472] P. J. Schmid. Dynamic mode decomposition for numerical and experimental data. *J. Fluid. Mech.*, 656:5–28, 2010.
- [473] P. J. Schmid, L. Li, M. P. Juniper, and O. Pust. Applications of the dynamic mode decomposition. *Theoretical and Computational Fluid Dynamics*, 25(1-4):249–259, 2011.
- [474] P. J. Schmid and J. Sesterhenn. Dynamic mode decomposition of numerical and experimental data. In *61st Annual Meeting of the APS Division of Fluid Dynamics*. American Physical Society, November 2008.
- [475] P. J. Schmid, D. Violato, and F. Scarano. Decomposition of time-resolved tomographic PIV. *Experiments in Fluids*, 52:1567–1579, 2012.
- [476] E. Schmidt. Zur theorie der linearen und nichtlinearen integralgleichungen. i teil. entwicklung willkürlichen funktionen nach system vorgeschriebener. *Math. Ann.*, 3:433–476, 1907.

- [477] Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *Science*, 324(5923):81–85, 2009.
- [478] Michael D. Schmidt, Ravishankar R. Vallabhajosyula, Jerry W. Jenkins, Jonathan E. Hood, Abhishek S. Soni, John P. Wikswo, and Hod Lipson. Automated refinement and inference of analytical models for metabolic networks. *Physical Biology*, 8(5):055011, 2011.
- [479] Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and beyond*. MIT press, 2002.
- [480] Gideon Schwarz et al. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.
- [481] A. Seena and H. J. Sung. Dynamic mode decomposition of turbulent cavity flows for self-sustained oscillations. *International Journal of Heat and Fluid Flow*, 32(6):1098–1110, 2011.
- [482] Ervin Sejdic', Igor Djurovic', and Jin Jiang. Time–frequency feature representation using energy concentration: An overview of recent advances. *Digital Signal Processing*, 19(1):153–183, 2009.
- [483] O. Semeraro, G. Bellani, and F. Lundell. Analysis of time-resolved PIV measurements of a confined turbulent jet using POD and Koopman modes. *Experiments in Fluids*, 53(5):1203–1220, 2012.
- [484] Onofrio Semeraro, Francois Lusseyran, Luc Pastur, and Peter Jordan. Qualitative dynamics of wavepackets in turbulent jets. *Physical Review Fluids*, 2(9):094605, 2017.
- [485] Gil Shabat, Yaniv Shmueli, Yariv Aizenbud, and Amir Averbuch. Randomized LU decomposition. *Applied and Computational Harmonic Analysis*, 2016.
- [486] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.
- [487] Ati S. Sharma, Igor Mezic', and Beverley J. McKeon. Correspondence between Koopman mode decomposition, resolvent mode decomposition, and invariant solutions of the Navier-Stokes equations. *Physical Review Fluids*, 1(3):032402, 2016.
- [488] Eli Shlizerman, Edwin Ding, Matthew O. Williams, and J. Nathan Kutz. The proper orthogonal decomposition for dimensionality reduction in mode-locked lasers and optical systems. *International Journal of Optics*, 2012, 2011.
- [489] Valeria Simoncini. A new iterative method for solving large-scale Lyapunov matrix equations. *SIAM Journal on Scientific Computing*, 29(3):1268–1288, 2007.
- [490] L. Sirovich. Turbulence and the dynamics of coherent structures, parts I-III. *Q. Appl. Math.*, XLV(3):561–590, 1987.
- [491] L. Sirovich and M. Kirby. A low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4(3):519–524, 1987.
- [492] S. Skogestad and I. Postlethwaite. *Multivariable Feedback Control*. Wiley, Chichester, 1996.
- [493] Paul Smolensky. Information processing in dynamical systems: Foundations of harmony theory. Technical report, Colorado Univ at Boulder Dept of Computer Science, 1986.

- [494] Giovanni Solari, Luigi Carassale, and Federica Tubino. Proper orthogonal decomposition in wind engineering. part 1: A state-of-the-art and some prospects. *Wind and Structures*, 10(2):153–176, 2007.
- [495] G. Song, F. Alizard, J.-C. Robinet, and X. Gloerfelt. Global and Koopman modes analysis of sound generation in mixing layers. *Physics of Fluids (1994-present)*, 25(12):124101, 2013.
- [496] Danny C. Sorensen and Yunkai Zhou. Direct methods for matrix Sylvester and Lyapunov equations. *Journal of Applied Mathematics*, 2003(6):277–303, 2003.
- [497] Mariia Sorokina, Stylianos Sygletos, and Sergei Turitsyn. Sparse identification for nonlinear optical communication systems: SINO method. *Optics Express*, 24(26):30433–30443, 2016.
- [498] James C. Spall. The Kantorovich inequality for error analysis of the Kalman filter with unknown noise distributions. *Automatica*, 31(10):1513–1517, 1995.
- [499] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.
- [500] W-H. Steeb and F. Wilhelm. Non-linear autonomous systems of differential equations and Carleman linearization procedure. *Journal of Mathematical Analysis and Applications*, 77(2):601–611, 1980.
- [501] Robert F. Stengel. *Optimal Control and Estimation*. Courier Corporation, 2012.
- [502] Gilbert W. Stewart. On the early history of the singular value decomposition. *SIAM Review*, 35(4):551–566, 1993.
- [503] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. Detecting causality in complex ecosystems. *Science*, 338(6106):496–500, 2012.
- [504] A. Surana. Koopman operator based observer synthesis for control-affine nonlinear systems. In *55th IEEE Conference on Decision and Control (CDC)*, pages 6492–6499, 2016.
- [505] Amit Surana and Andrzej Banaszuk. Linear observer synthesis for nonlinear systems using Koopman operator framework. *IFAC-PapersOnLine*, 49(18):716–723, 2016.
- [506] Yoshihiko Susuki and Igor Mezic. A prony approximation of Koopman mode decomposition. In *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*, pages 7022–7027. IEEE, 2015.
- [507] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*, volume 1. MIT press Cambridge, 1998.
- [508] Adam Svenkeson, Bryan Glaz, Samuel Stanton, and Bruce J. West. Spectral decomposition of nonlinear systems with memory. *Phys. Rev. E*, 93:022211, Feb 2016.
- [509] S. A. Svoronos, D. Papageorgiou, and C. Tsiligiannis. Discretization of nonlinear control systems via the Carleman linearization. *Chemical Engineering Science*, 49(19):3263–3267, 1994.
- [510] D. L. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 18(8):831–836, 1996.

- [511] K. Taira and T. Colonius. The immersed boundary method: a projection approach. *Journal of Computational Physics*, 225(2):2118–2137, 2007.
- [512] Naoya Takeishi, Yoshinobu Kawahara, Yasuo Tabei, and Takehisa Yairi. Bayesian dynamic mode decomposition. *Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2017.
- [513] Naoya Takeishi, Yoshinobu Kawahara, and Takehisa Yairi. Learning Koopman invariant subspaces for dynamic mode decomposition. In *Advances in Neural Information Processing Systems*, pages 1130–1140, 2017.
- [514] Naoya Takeishi, Yoshinobu Kawahara, and Takehisa Yairi. Subspace dynamic mode decomposition for stochastic Koopman analysis. *Physical Review*, E 96.3:033310, 2017.
- [515] F. Takens. Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*, 898:366–381, 1981.
- [516] Z. Q. Tang and N. Jiang. Dynamic mode decomposition of hairpin vortices generated by a hemisphere protuberance. *Science China Physics, Mechanics and Astronomy*, 55(1):118–124, 2012.
- [517] Roy Taylor, J. Nathan Kutz, Kyle Morgan, and Brian Nelson. Dynamic mode decomposition for plasma diagnostics and validation. *Review of Scientific Instruments*, 89 (5):053501, 2018.
- [518] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.
- [519] Zhou Ting and Jiang Hui. Eeg signal processing based on proper orthogonal decomposition. In *Audio, Language and Image Processing (ICALIP), 2012 International Conference on*, pages 636–640. IEEE, 2012.
- [520] Santosh Tirunagari, Norman Poh, Kevin Wells, Mirosław Bober, Isky Gorden, and David Windridge. Movement correction in DCE-MRI through windowed and reconstruction dynamic mode decomposition. *Machine Vision and Applications*, 28(3-4):393–407, 2017.
- [521] Christopher Torrence and Gilbert P. Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1):61–78, 1998.
- [522] Giang Tran and Rachel Ward. Exact recovery of chaotic systems from highly corrupted data. *Multiscale Modeling & Simulation*, 15 (3):1108–1129, 2017.
- [523] Lloyd N. Trefethen. *Spectral methods in MATLAB*. SIAM, 2000.
- [524] Lloyd N. Trefethen and David Bau III. *Numerical linear algebra*, volume 50. SIAM, 1997.
- [525] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231–2242, 2004.
- [526] J. A. Tropp. Recovery of short, complex linear combinations via l_1 minimization. *IEEE Transactions on Information Theory*, 51(4):1568–1570, 2005.
- [527] J. A. Tropp. Algorithms for simultaneous sparse approximation. part ii: Convex relaxation. *Signal Processing*, 86(3):589–602, 2006.
- [528] J. A. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, 52(3):1030–1051, 2006.
- [529] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, 2007.

- [530] J. A. Tropp, A. C. Gilbert, and M. J. Strauss. Algorithms for simultaneous sparse approximation. part i: Greedy pursuit. *Signal Processing*, 86(3):572–588, 2006.
- [531] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk. Beyond Nyquist: Efficient sampling of sparse bandlimited signals. *IEEE Transactions on Information Theory*, 56(1):520–544, 2010.
- [532] Joel A. Tropp, Alp Yurtsever, Madeleine Udell, and Volkan Cevher. Randomized single-view algorithms for low-rank matrix approximation. *arXiv preprint arXiv:1609.00048*, 2016.
- [533] J. H. Tu and C. W. Rowley. An improved algorithm for balanced POD through an analytic treatment of impulse response tails. *J. Comp. Phys.*, 231(16):5317–5333, 2012.
- [534] J. H. Tu, C. W. Rowley, E. Aram, and R. Mittal. Koopman spectral analysis of separated flow over a finite-thickness flat plate with elliptical leading edge. *AIAA Paper 2011*, 2864, 2011.
- [535] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz. On dynamic mode decomposition: theory and applications. *J. Comp. Dyn.*, 1(2):391–421, 2014.
- [536] Jonathan H. Tu, Clarence W. Rowley, J. Nathan Kutz, and Jessica K. Shang. Spectral analysis of fluid flows using sub-Nyquist-rate PIV data. *Experiments in Fluids*, 55(9):1–13, 2014.
- [537] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.
- [538] Rudolph Van Der Merwe. *Sigma-point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models*. 2004.
- [539] Charles Van Loan. *Computational Frameworks for the Fast Fourier Transform*. SIAM, 1992.
- [540] Daniele Venturi and George Em Karniadakis. Gappy data and reconstruction procedures for flow past a cylinder. *Journal of Fluid Mechanics*, 519:315–336, 2004.
- [541] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*, pages 1096–1103. ACM, 2008.
- [542] Stefan Volkwein. Model reduction using proper orthogonal decomposition. *Lecture Notes, Institute of Mathematics and Scientific Computing, University of Graz*. see <http://www.uni-graz.at/imawww/volkwein/POD.pdf>, 1025, 2011.
- [543] Stefan Volkwein. Proper orthogonal decomposition: Theory and reduced-order modelling. *Lecture Notes, University of Konstanz*, 4:4, 2013.
- [544] Sergey Voronin and Per-Gunnar Martinsson. RSVDPACK: Subroutines for computing partial singular value decompositions via randomized sampling on single core, multi core, and GPU architectures. *arXiv preprint arXiv:1502.05366*, 2015.
- [545] Avery Wang et al. An industrial strength audio search algorithm. In *Ismir*, volume 2003, pages 7–13. Washington, DC, 2003.
- [546] H. H. Wang, M. Krstić, and G. Bastin. Optimizing bioreactors by extremum seeking. *Adaptive Control and Signal Processing*, 13(8):651–669, 1999.

- [547] H. H. Wang, S. Yeung, and M. Krstic'. Experimental application of extremum seeking on an axial-flow compressor. *IEEE Transactions on Control Systems Technology*, 8(2):300–309, 2000.
- [548] W. X. Wang, R. Yang, Y. C. Lai, V. Kovanis, and C. Grebogi. Predicting catastrophes in nonlinear dynamical systems by compressive sensing. *Physical Review Letters*, 106:154101–1–154101–4, 2011.
- [549] Zhu Wang, Imran Akhtar, Jeff Borggaard, and Traian Iliescu. Proper orthogonal decomposition closure models for turbulent flows: a numerical comparison. *Computer Methods in Applied Mechanics and Engineering*, 237:10–26, 2012.
- [550] Christoph Wehmeyer and Frank Noé. Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *The Journal of Chemical Physics*, 148(24):241703, 2018.
- [551] Greg Welch and Gary Bishop. An introduction to the Kalman filter, 1995.
- [552] Peter Whittle. *Hypothesis Testing in Time Series Analysis*, volume 4. Almqvist & Wiksells, 1951.
- [553] O. Wiederhold, R. King, B. R. Noack, L. Neuhaus, L. Neise, W. an Enghard, and M. Swoboda. Extensions of extremum-seeking control to improve the aerodynamic performance of axial turbomachines. In *39th AIAA Fluid Dynamics Conference*, pages 1–19, San Antonio, TX, USA, 2009. AIAA-Paper 092407.
- [554] K. Willcox and J. Peraire. Balanced model reduction via the proper orthogonal decomposition. *AIAA Journal*, 40(11):2323–2330, 2002.
- [555] Karen Willcox. Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition. *Computers & Fluids*, 35(2):208–226, 2006.
- [556] Matthew O. Williams, Ioannis G. Kevrekidis, and Clarence W. Rowley. A data-driven approximation of the Koopman operator: extending dynamic mode decomposition. *Journal of Nonlinear Science*, 6:1307–1346, 2015.
- [557] Matthew O. Williams, Clarence W. Rowley, and Ioannis G. Kevrekidis. A kernel approach to data-driven Koopman spectral analysis. *Journal of Computational Dynamics*, 2(2):247–265, 2015.
- [558] Daniela M. Witten and Robert Tibshirani. Penalized classification using Fisher's linear discriminant. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(5):753–772, 2011.
- [559] Franco Woolfe, Edo Liberty, Vladimir Rokhlin, and Mark Tygert. A fast randomized algorithm for the approximation of matrices. *Journal of Applied and Computational Harmonic Analysis*, 25:335–366, 2008.
- [560] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(2):210–227, 2009.
- [561] C. F. Jeff Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, pages 95–103, 1983.
- [562] Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, S. Yu Philip et al. Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1):1–37, 2008.
- [563] Hao Ye, Richard J. Beamish, Sarah M. Glaser, Sue C. H. Grant, Chih-hao Hsieh, Laura J. Richards, Jon T. Schnute, and George Sugihara. Equation-

- free mechanistic ecosystem forecasting using empirical dynamic modeling. *Proceedings of the National Academy of Sciences*, 112(13):E1569–E1576, 2015.
- [564] Enoch Yeung, Soumya Kundu, and Nathan Hodas. Learning deep neural network representations for Koopman operators of nonlinear dynamical systems. *arXiv preprint arXiv:1708.06850*, 2017.
- [565] B. Yildirim, C. Chrysostomidis, and G. E. Karniadakis. Efficient sensor placement for ocean measurements using low-dimensional concepts. *Ocean Modelling*, 27(3):160–173, 2009.
- [566] Xiaoming Yuan and Junfeng Yang. Sparse and low-rank matrix decomposition via alternating direction methods. *preprint*, 12, 2009.
- [567] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus. Deconvolutional networks. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 2528–2535, 2010.
- [568] C. Zhang and R. Ordóñez. Numerical optimization-based extremum seeking control with application to ABS design. *IEEE Transactions on Automatic Control*, 52(3):454–467, 2007.
- [569] Hao Zhang, Clarence W. Rowley, Eric A. Deem, and Louis N. Cattafesta. Online dynamic mode decomposition for time-varying systems. *arXiv preprint arXiv:1707.02876*, 2017.
- [570] T. Zhang, G. Kahn, S. Levine, and P. Abbeel. Learning deep control policies for autonomous aerial vehicles with MPC-guided policy search. In *IEEE Robotics and Automation (ICRA)*, pages 528–535, 2016.
- [571] Weiwei Zhang, Bobin Wang, Zhengyin Ye, and Jingge Quan. Efficient method for limit cycle flutter analysis based on nonlinear aerodynamic reduced-order models. *AIAA Journal*, 50(5):1019–1028, 2012.
- [572] Sanjo Zlobec. An explicit form of the moore-penrose inverse of an arbitrary complex matrix. *SIAM Review*, 12(1):132–134, 1970.
- [573] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.