Electrochemical Model-Based State of Charge and State of Health Estimation of Lithium-Ion Batteries

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the Graduate School of The Ohio State University

By

Alexander Paul Bartlett, M.S.

Graduate Program in Mechanical Engineering

* * * * *

The Ohio State University

2015

Dissertation Committee

Giorgio Rizzoni, Advisor
Marcello Canova
Terry Conlisk
Yann Guezennece
Vehicle electrification continues to be a key topic of interest for automotive manufacturers, in an effort to reduce the usage of fossil fuel energy and improve vehicle efficiency. Lithium-ion batteries are currently the technology of choice for hybrid and electric vehicles due to their decreasing cost and improved power and energy density over traditional lead-acid or nickel-metal-hydride batteries. In particular, batteries with composite electrodes have seen increased use in automotive applications due to their ability to balance energy density, power density, and cost by adjusting the amount of each material within the electrode. However, this improved performance introduces new challenges to ensure the battery pack operates safely, reliably, and durably. The vehicle’s battery management system (BMS) is designed to meet these challenges, in part, by estimating the battery state of charge (SOC) and state of health (SOH). Knowledge of SOC allows the BMS to predict the available instantaneous power while ensuring the battery is operating within safe limits. As batteries age, they lose capacity and the ability to deliver power. Therefore, tracking the battery SOH is necessary to maintain accurate estimates of SOC and power throughout the battery life and give an accurate miles-to-empty metric to the driver.

Recently, increased attention has been given to electrochemical models for SOC and SOH estimation, over traditional circuit models. Electrochemical models based on first principles have the potential to more accurately predict cell performance, as well as provide more information about the internal battery states. State of health estimation algorithms that do not use electrochemical-based models may have more
difficulty maintaining an accurate battery model as the cell ages under varying degra-
dation modes such as lithium consumption at the solid-electrolyte interface or active material dissolution. However, efforts to validate electrochemical model-based state estimation algorithms with experimental aging data are limited, particularly for composite electrode cells.

This dissertation focuses on applying electrochemical models towards SOC and SOH estimation problems, specifically on automotive cells with a composite LiMn$_2$O$_4$-LiNi$_{1/3}$Mn$_{1/3}$Co$_{1/3}$O$_2$ (LMO-NMC) cathode. Model order reduction techniques are applied to reduce the computational complexity of the governing electrochemical equations, while still maintaining physically meaningful parameters. The observability of the reduced order model is analyzed in order to facilitate state estimation, while also determining the limitations on the number of capacity and power related parameters that can be estimated simultaneously. Various nonlinear observers are implemented to estimate the model states associated with SOC and SOH, in order to determine their applicability in an onboard battery management system. Finally, SOC and SOH estimates are validated with experimental aging data.
Acknowledgements

I would like to take the time to properly thank everyone that contributed to this work, as I could not have done it alone.

I would like to thank my advisor, Dr. Rizzoni, for giving me this incredible opportunity here at CAR, and for all of the advice over the years. A special thanks is due for always providing excellent coffee at our meetings and for the golf tips, including helping me learn not to over-swing.

I would also like to thank Dr. Canova, Dr. Conlisk, and Dr. Guezenne for serving on my dissertation committee, and for their assistance with the modeling work and experiments.

I am especially grateful to my former fellow grad student Jim Marcicki for being an amazingly patient and helpful mentor, always taking the time to help me understand the world of battery modeling, and continuing to offer advice and suggestions throughout my graduate career.

I would like to thank my colleges at CAR, especially Guodong, Ke, Samartha, Shrikant, Meng, Chuting, and Dell for all of our enlightening conversations about various battery and non-battery related topics. Also, special thanks to Nick Warner for setting up and running the aging experiments and to Dr. Co and my friend Jon Goiri for patiently trying to explain chemistry to this clueless mechanical engineer.

I want to thank Ford Motor Company for their support via the Ford University Research Program (URP), as well as for the opportunities for two summer internships.
A special thanks to my colleges at Ford for all of their help, including Ted Miller, Xiao Guang Yang, Kevin Rhodes, and OuJung Kwon.

Finally, and most importantly, I would like to thank my family for everything they have done for me. A special thanks to my Mom and Dad for their love, support, advice, and constant reminders not to work too hard, my brother Andy and sister-in-law Michele for making these four years a more cromulent experience, and my Grandparents for their constant cheering from the sidelines. Also, thanks to my friend Peter for helping me keep my sanity, and to the friendly staff at Reno’s East. Thanks for finally getting some decent darts.
Vita

May 20, 1988 ....................................... Born - Columbus, Ohio

2011 ..................................................... B.S. Mechanical Engineering, Summa Cum Laude, University of Michigan

2014 ..................................................... M.S. Mechanical Engineering, The Ohio State University

2014 ..................................................... Graduate Student, The Ohio State University

Publications


Fields of Study

Major Field: Mechanical Engineering
# Table of Contents

Abstract ................................................................. ii
Acknowledgements ......................................................... iv
Vita ................................................................. vi
Table of Contents ......................................................... vii
List of Tables ........................................................... x
List of Figures ............................................................. xi

1 Introduction and Overview ............................................. 1
   1.1 An Introduction to Vehicle Electrification ....................... 2
   1.2 Batteries for Energy Storage ...................................... 3
   1.3 Types of Batteries ................................................ 6
   1.4 Basic Function of Batteries ...................................... 7
   1.5 Battery Electrodes ................................................ 9
   1.6 Aging Mechanisms ................................................ 13
   1.7 Battery Management System .................................... 15
   1.8 Basic Terminology ............................................... 17
   1.9 Thesis Objectives ............................................... 19

2 Battery Modeling ..................................................... 21
   2.1 Introduction .................................................... 21
   2.2 A Hierarchy of Battery Models ................................ 22
   2.3 Nomenclature ................................................... 24
   2.4 The Porous Electrode Model .................................... 26
      2.4.1 1+1D porous electrode model ................................ 28
      2.4.2 Governing electrochemical equations ...................... 29
      2.4.3 The open circuit voltage from thermodynamics ........... 29
      2.4.4 Kinetic overpotential ...................................... 33
      2.4.5 Mass transport in the spherical electrode particle and the concentration overpotential .......................... 40
      2.4.6 Electrical conduction in the solid ......................... 42
      2.4.7 Lithium transport in the electrolyte ....................... 42
      2.4.8 Potential in the liquid ..................................... 50
      2.4.9 Solution method ............................................ 52
   2.5 The Single Particle Model ...................................... 53
   2.6 Model Order Reduction .......................................... 54
3 Methods for State Estimation
3.1 Introduction to Estimation 84
3.2 Observability 86
  3.2.1 Observability for linear systems 87
  3.2.2 Observability for nonlinear systems 91
  3.2.3 Observability of the single particle model 95
3.3 Observer Design 99
  3.3.1 The extended Kalman filter 101
  3.3.2 The particle filter 108
  3.3.3 Other nonlinear observers 113
3.4 Methods for Smoothing 115
3.5 Combined State and Parameter Estimation 121
3.6 Summary 124

4 State of Charge Estimation 126
4.1 Introduction 126
  4.1.1 Defining state of charge 127
  4.1.2 Methods for state of charge estimation 129
4.2 SOC Estimation for a Composite Electrode 130
  4.2.1 Experimental methods 131
  4.2.2 Calculating the current split 133
  4.2.3 Calculating a cell level SOC 143
  4.2.4 Initializing the electrodes 147
4.3 Observability of the Complete Model and Possible Improvements 148
  4.3.1 Approach 1: Open-loop estimation of one electrode 151
  4.3.2 Approach 2: Steady-state approximation for one electrode 153
  4.3.3 Observability of open-loop and steady-state methods 156
4.4 Validation of SOC Estimates 162
4.5 Summary and Future Work 166

5 State of Health Estimation 169
5.1 Introduction 169
  5.1.1 Defining state of health 170
  5.1.2 Methods for state of health estimation 171
5.2 Estimating Active Material Loss 172
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2.1</td>
<td>Observer design</td>
<td>175</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Experimental validation of active material estimates</td>
<td>179</td>
</tr>
<tr>
<td>5.3</td>
<td>Estimating Loss of Cyclable Lithium</td>
<td>193</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Validation with experimental data</td>
<td>202</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Quantifying SOC/SOH Estimate Error</td>
<td>212</td>
</tr>
<tr>
<td>5.4</td>
<td>Simultaneous Estimation of Degradation Mechanisms</td>
<td>226</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Combined active material/cyclable lithium estimation algorithm</td>
<td>228</td>
</tr>
<tr>
<td>5.5</td>
<td>Summary and Future Work</td>
<td>234</td>
</tr>
</tbody>
</table>

### 6 Summary

#### A Observability Calculations

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1</td>
<td>Observability of the 2-Electrode Single Particle Model</td>
<td>249</td>
</tr>
</tbody>
</table>
# List of Tables

1.1 Standard reference potential and theoretical specific capacity for different electrode materials [49,100]. ...................................................... 11

2.1 Linear transfer function coefficients for solid diffusion using a Padé approximation of various orders. ...................................................... 61

2.2 Linear transfer function coefficients for liquid diffusion using a Padé approximation of various orders. ...................................................... 67

3.1 Extended Kalman filter algorithm for state estimation of a generic nonlinear system with assumed process and sensor noise. ............ 105

3.2 Unscented Kalman filter algorithm for a generic nonlinear system with assumed process and sensor noise. ...................................................... 115

3.3 Kalman smoother algorithm for state estimation of a generic nonlinear system with assumed process and sensor noise. .................. 120

4.1 Model parameters used for the composite LMO/NMC electrode. ... 135

4.2 Model parameters used for the negative electrode and electrolyte. ... 136

4.3 Model equations for estimating the states of both electrodes, estimating the states of one electrode using the open-loop method, or estimating the positive states and using the steady-state approximation to obtain the negative concentration. The OCVs in the output can be replaced by a cubic spline section, as in Section 3.2.3, to facilitate calculation of the Lie derivatives. ...................................................... 157

5.1 Test conditions for the different cells tested. ................................. 202

5.2 Capacity Estimation Results ...................................................... 205
# List of Figures

1.1 Ragone plot of energy density versus power density for various vehicle energy storage technologies [92]. ......................................................... 3
1.2 Energy density and costs trends of commercially available Li-ion batteries for PHEV applications [2]. .......................................................... 5
1.3 Schematic showing the operation of a Li-ion battery. ......................... 8

2.1 A hierarchy of battery models. .......................................................... 22
2.2 A schematic of the 1+1D model structure. The $x$ direction spans through-the-thickness of the cell from one electrode end to the other, while the $r$ direction spans the radius of each particle from 0 to the particle radius. .......................................................... 28
2.3 Open circuit voltage versus Li concentration of a graphite electrode. 34
2.4 Schematic of the change in Gibbs free energy going from reactants to products as a function of the reaction coordinate. ......................... 35
2.5 Schematic of the concentration gradient that forms as the redox reactions occur at the electrode particle surface. ................................. 40
2.6 Schematic showing the differences in chemical potential between region 1 and 2, and the opportunity to lower the Gibbs free energy via diffusion. 45
2.7 Schematic showing the differences in electrostatic potential between position 1 and 2. Particles will move toward regions of lower electrochemical potential. ......................................................... 46
2.8 Representative concentration profiles along the radial distance $r$ at various times during a 1C discharge. ................................................. 57
2.9 Step response of the surface concentration for various reduced order models, compared to the original PDE numerical solution. .......... 62
2.10 Frequency response of the surface concentration for various reduced order models, compared to the original PDE transcendental transfer function. ......................................................... 63
2.11 Representative liquid concentration profiles along the $x$ direction, at various times for a 1C discharge. ................................................. 65
2.12 Step response of the liquid concentration for various Padé approxima-
tion orders, compared to the original PDE numerical solution. .......... 68
2.13 Frequency response of the liquid concentration at the negative electrode for various Pade approximation orders, compared to the original transcendental transfer function. .................................................. 69
2.14 Schematic of the electrochemical model, with a single negative electrode particle and two positive electrode particles acting in parallel. ................................................................. 70
2.15 Schematic of the spatial variation in the current carried in the solid and liquid phases, under the uniform utilization assumption. ................................................................. 76
2.16 Model prediction of a 1C discharge curve under different degradation mechanisms, each resulting in 20% capacity fade. ................................................................. 81

3.1 Typical structure of a model-based observer, involving prediction, output residual generation, and correction. ................................................................. 101
3.2 Example of the effect tuning the process noise covariance, $Q$, has on the state estimate. ................................................................. 106
3.3 Example of the effect tuning the sensor noise covariance, $R$, has on the state estimate. ................................................................. 107
3.4 Visualization of the particle filter algorithm. ................................................................. 112
3.5 Example showing forward, backward, and smoothed estimates of normalized concentration using the single particle model to generate pseudo-data. ................................................................. 119
3.6 Single particle example showing the trace of the estimate covariance matrices for forward, backward, and smoothed estimates. ................................................................. 121
3.7 Forward, backward, and smoothed normalized concentration estimates when a 50% error in the electrode area is introduced. ................................................................. 122

4.1 Charge-depleting profile defined by the USABC for a PHEV application, used to cycle the 15 Ah pouch cells. ................................................................. 132
4.2 Measured voltage and predicted voltage during a 1C constant current discharge. ................................................................. 134
4.3 Measured voltage and predicted voltage for the charge depleting current profile used in the aging campaign. ................................................................. 134
4.4 Comparison of the calculated current split for each composite material, between the iterative Newton-Raphson method and the non-iterative linearization method. Subplot b.) shows a close-up of the initial pulses. 140
4.5 Comparison of the calculated current going to each composite material particle, between the iterative Newton-Raphson method and the non-iterative linearization method. Subplot b.) shows a close-up of the initial pulses. ................................................................. 141
4.6 Definition of the electrode composition ranges corresponding to the upper and lower cell voltage limit. ................................................................. 144
4.7 LMO and NMC OCVs as a function of normalized concentration. Initial and final SOC are defined at the upper and lower cell voltage limits. The overall initial and final SOC, $y_0$ and $y_f$ correspond to a cell SOC (top axis) of 1 and 0, respectively. ................................................................. 146
4.8 Surface SOC estimate results when attempting to estimate the states of both electrodes using an EKF. A perfect model with added sensor noise is assumed in order to generate pseudo-data, and an initial condition error on the positive states is imposed.

4.9 Block diagram of the “open-loop” algorithm for estimating the positive electrode states. The negative electrode and liquid phase are simulated open-loop in order to generate a pseudo-measurement of the positive electrode potential. Alternatively, the negative electrode states could be estimated by simulating the positive electrode open-loop.

4.10 Surface SOC estimate results when using the open-loop method, where the states of only the positive electrode are estimated using an EKF. A perfect model with added sensor noise is assumed in order to generate pseudo-data, and an initial condition error on the positive states is imposed.

4.11 Surface SOC estimate results when using the open-loop method, where the states of only the negative electrode are estimated using an EKF. A perfect model with added sensor noise is assumed in order to generate pseudo-data, and an initial condition error on the negative states is imposed.

4.12 Block diagram of the “steady-state” algorithm for estimating the positive electrode states. The negative electrode bulk concentration is obtained from the positive electrode bulk concentration via conservation of lithium.

4.13 Surface SOC estimate results when using the steady-state method, where the positive electrode states are estimated using an EKF, and the negative electrode bulk concentration is obtained via conservation of lithium. A perfect model with added sensor noise is assumed in order to generate pseudo-data, and an initial condition error on the positive states is imposed.

4.14 Log of the observability matrix condition number, plotted over the course of a CD-CCCV profile. In a.) the states of both electrodes are assumed to be estimated, in b.) the states of only one electrode are estimated (the other is simulated open-loop), and in c.) the positive electrode states are estimated, with the negative electrode bulk concentration obtained via the steady-state method.

4.15 Estimation results over a CD, CCCV charge profile. a) Imposed current profile, b) predicted and measured cell voltage, c) cell SOC estimate and SOC from Coulomb counting, showing convergence from an initial condition error, and d) cell SOC estimate error.

4.16 Estimation results for cell 2 over a CD, CCCV charge profile. a) Estimated bulk and surface normalized concentrations for each composite material, b) current split factor, \( \beta_{LMO} \) \( \beta_{NMC} \) is simply equal to \( 1 - \beta_{LMO} \), and c) the current (in A) going to the LMO particle compared to the total current.
4.17 Cell SOC estimate error, comparing open-loop and steady-state methods.
4.18 Cell SOC estimate error, comparing open-loop and steady-state methods, with a 15% error imposed on the $x_0$ parameter.

5.1 Log of the observability matrix condition number in the case of estimating active material volume fractions, over the course of a CD-CCCV profile.
5.2 1 cycle of the CD-CS-CCCV current profile (a.) used to generate pseudo-data, and the resulting voltage response (b.).
5.3 Volume fraction estimate results showing a.) the total active material volume fractions, b.) the overall active material volume fraction, and c.) the volume fraction splits.
5.4 Open loop model fit of a C/5 discharge at BOL for the a.) 1:1 ratio and b.) 3:1 ratio cells.
5.5 Active material volume fraction estimation results for the 1:1 ratio cell.
5.6 Active material volume fraction estimation results for the 3:1 ratio cell.
5.7 Model-generated lookup table used to convert volume fraction estimates to a cell capacity for the 1:1 ratio cell (3:1 ratio cell not shown). Lines on the figure represent possible combinations of $\epsilon_{LMO}$ and $\epsilon_{NMC}$ for a given cell capacity. The two points represent the specific combination of $\epsilon_{LMO}$ and $\epsilon_{NMC}$ estimated by the EKF at the beginning and end of the experiment.
5.8 Estimated cell capacity based on volume fraction estimates for the a.) 1:1 ratio and b.) 3:1 ratio cells, compared to the measured capacity.
5.9 Experimental differential capacity curves measured at various stages of life for the a.) 1:1 ratio and b.) 3:1 ratio cells.
5.10 Model-generated differential capacity curves under different aging scenarios: a.) loss of LMO material, b.) loss of NMC material, and c.) loss of overall active material. Each scenario corresponds to a signature change in the peaks at various voltages.
5.11 Experimental dQ/dV peaks versus measured cell capacity for a.) 1:1 ratio and b.) 3:1 ratio cells, showing a roughly linear relationship for each peak, following the initial 3 cycles.
5.12 Example of a shift in the electrode operating range from a new to an aged cell, as cyclable lithium is lost to SEI growth. In each electrode, the operating range tends to shift to a less lithiated state.
5.13 Visualization of the decrease in electrode compositions $x_f$ and $y_f$ resulting from a loss in lithium over a given trip. A similar diagram could be drawn for a charge to see the decrease in $x_0$ and $y_0$. In this example, assume voltage limits of 2.8-4.1 V.
5.14 Demonstration of the increased negative electrode concentration estimate error in regions where the OCV is flat. A perfect model is assumed in order to generate pseudo-data. A constant 20 mV voltage measurement bias is applied to the pseudo-data voltage. The pseudo-data is then used as the “measurement” in the EKF to estimate the negative electrode concentration.

5.15 Negative electrode OCV (measured from half cell experiment), showing regions where the OCV is flat or steep.

5.16 Block diagram of the combined cell SOC/SOH estimation algorithm using the open-loop method. Cell SOC is estimated continuously from estimates of the positive electrode concentrations, while a capacity estimate is obtained once every CD cycle from an estimate of the negative electrode concentration range $x_0$ to $x_f$.

5.17 Capacity estimation results using the open-loop method for Cell 1 aged, with a CD profile to 25% SOC, charged at C/3.

5.18 Capacity estimation results using the open-loop method for Cell 2, aged with a CD profile to 25% SOC, charged at 3C/2.

5.19 Capacity estimation results using the open-loop method for Cell 3, aged with a CD profile to 25% SOC, charged at 5C.

5.20 Capacity estimation results using the open-loop method for Cell 4, aged with a CD profile to 35% SOC, charged at C/3.

5.21 Capacity estimation results using the open-loop method for Cell 5, aged with a CD profile to 45% SOC, charged at C/3.

5.22 Capacity estimation results for Cell 1, comparing the open-loop and steady-state methods.

5.23 Capacity estimation results for Cell 2, comparing the open-loop and steady-state methods.

5.24 Capacity estimation results for Cell 3, comparing the open-loop and steady-state methods.

5.25 Capacity estimation results for Cell 4, comparing the open-loop and steady-state methods.

5.26 Capacity estimation results for Cell 5, comparing the open-loop and steady-state methods.

5.27 Root mean square and maximum SOC estimate error for cell 2. Each point indicates the error over a single CD-CCCV profile.

5.28 Diagram of the propagation of estimate error for SOC and capacity estimates.

5.29 Absolute value of the difference between predicted and measured cell voltage, over the course of a CD cycle.

5.30 $SOC_n$ estimate results using model generated pseudo-data with a constant 14 mV voltage bias applied in either direction.

5.31 Error in the $x_0$ estimate during a CD cycle, in response to an error in the initial assumed $x_0$. Here, a perfect model is assumed.
5.32 Error in the $x_0$ estimate during a CD cycle, in response to an error in the initial assumed $x_0$. Here a 14 mV voltage bias is applied to the model-generated pseudo-data. The shaded areas represent all possible results for estimates taken at any point within the mid-SOC or low-SOC regions.

5.33 Capacity estimation results for cell 1, with dashed lines indicated the computed error bounds of ± 0.53 Ah, applied to the EKF estimates.

5.34 Capacity estimation results for cell 5, with dashed lines indicated the computed error bounds of ± 2.0 Ah, applied to the EKF estimates.

5.35 Cell SOC error in response to a.) a ± 14 mV voltage bias, b.) a 7% or 16% error in $x_0$, corresponding to $x_0$ estimates taken at low-SOC or mid-SOC, respectively, and c.) both voltage bias and $x_0$ error.

5.36 Comparison of measured and estimated cell SOC, with bounds shown for both the measurement and the estimate.

5.37 Model error over the course of the CD profile, defined as the difference between the measured cell voltage and the model prediction. This dynamic error profile is then imposed on the voltage pseudo-data to replicate experimental data.

5.38 $x_0$ estimates under scenarios of a perfect model, dynamic model error, and voltage sensor bias.

5.39 Positive electrode active material volume fraction estimates under scenarios of a perfect model, dynamic model error, and voltage sensor bias.

5.40 $x_0$ estimates under scenarios of a perfect model, dynamic model error, and voltage sensor bias, assuming a reference electrode is available.

5.41 Positive electrode active material volume fraction estimates under scenarios of a perfect model, dynamic model error, and voltage sensor bias, assuming a reference electrode is available.
Chapter 1

Introduction and Overview

Several decades ago, hybrid and electric vehicles were mostly relegated to prototypes or expensive novelties. Since then, a rapid advancement in electric drivetrain technologies has facilitated the wide-spread commercialization of electrified vehicles. Today, electrified vehicles are mass produced in some form by nearly every major auto manufacturer, and continue to see an exponential-like growth in sales [2]. Advancement in battery technology is a major contributor to the success of electrified vehicles. In particular, the high energy and power density of lithium-ion batteries make them the current chemistry of choice in the automotive market. The inherent advantages of electrified vehicles over traditional vehicles, including the potential to significantly improve energy efficiency and reduce greenhouse gas emissions, should continue to drive their market penetration in the foreseeable future. However, adoption of these new technologies requires careful management to ensure that electrified vehicles operate safely and reliably.
1.1 An Introduction to Vehicle Electrification

Electrified vehicles can be broadly categorized into three different types: hybrid-electric vehicles (HEVs), plug-in hybrid-electric vehicles (PHEVs), and battery-electric vehicles (BEVs or EVs). Note that the acronym xEV is sometimes used to describe any electrified vehicle category, where the letters H, PH, or B can be substituted for x. In an HEV, the traditional internal combustion engine (ICE) powertrain is augmented with a battery pack and electric machine (EM) that are used to assist the engine in accelerating the vehicle and to recover energy during braking. The role of the battery in an HEV is mostly to assist the ICE in short bursts rather than to power the vehicle on its own; therefore, battery power is valued over battery capacity. HEVs operate in charge sustaining (CS) mode, meaning the battery is managed so that it is never fully depleted or fully charged, providing power when it can but not so much that the battery gets too depleted. The power capabilities of the battery pack and EM dictate how much the battery can assist the engine; the battery pack in a micro hybrid may only be able to start the engine rapidly from a standstill (facilitating engine “stop-start”), while the pack in a full hybrid may be able to power the entire vehicle on its own in short bursts.

A BEV eliminates the engine altogether and relies entirely on the battery pack to power the vehicle. The battery is charged by plugging into the grid, as well as through regenerative braking during operation. A BEV requires a large battery pack with a high capacity in order to deliver an acceptable range.

A PHEV is essentially a crossover between an HEV and a BEV, containing an engine and a moderately sized battery pack. A PHEV typically has an all-electric range of 10-40 miles when fully charged, during which the engine is either turned off or providing minimal power. In this all-electric range the battery is operating in charge
depleting (CD) mode, where more energy is being used to power the vehicle than is being recovered by braking, resulting in a gradual depletion of the battery. Once the battery is depleted to some minimum threshold, the vehicle switches to charge sustaining operation. In this way, a PHEV acts like a BEV throughout its initial all-electric range and like an HEV afterwards. However, this dual mode operation requires a battery pack that has both high power and high capacity, making it in many respects the most challenging application to design for.

1.2 Batteries for Energy Storage

Energy storage has a large number of applications, ranging from small-scale portable electronics to large-scale power generation. There are many different types of energy storage devices that store different types of energy. For example, springs, flywheels, pumped hydro, and compressed air systems store mechanical energy, capacitors store electrical energy, fossil fuels and hydrogen store chemical energy, and supercapacitors and batteries store electrochemical energy. Batteries in particular are useful for
many energy storage applications due to their scalability and balance between power and energy density. Individual battery cells just several millimeters long can power implantable medical devices, while on the other hand many thousands of cells can be arranged into grid storage units, storing megawatts of energy to be used in conjunction with intermittent renewable energy systems like wind and solar. In automotive applications, any energy storage system has to contend with limitations on space and weight; therefore, energy and power density per unit volume or mass become important metrics to consider.

Figure 1.1 shows a Ragone plot of energy density versus power density for different automotive energy storage technologies. Energy density is related to the distance the vehicle can travel using the battery pack, while power density is related to the vehicle’s acceleration. From the plot, it should be no surprise that gasoline and diesel powered internal combustion engines have dominated the automotive market throughout the 20th century, as they have an order of magnitude higher energy density than any of the commercial battery technologies. While battery energy density and cost continue to improve, as shown in Fig. 1.2, they are still two of the biggest liabilities limiting widespread adoption of xEVs. Another limiting factor is battery aging, as the battery’s ability to store energy and deliver power tends to decrease over time. This results in a diminishing all-electric range, efficiency, and acceleration performance as the battery ages.

Despite these shortcomings, there are advantages inherent to batteries that make them attractive either to complement or replace an IC engine. The most obvious reason to utilize onboard battery storage is to improve vehicle efficiency. First, a battery powered electric machine (EM) has the ability to replace the traditional starter and alternator, allowing the engine to be shut off while the vehicle is stopped at a light or in traffic. When acceleration is needed again, the battery pack can quickly start the engine and provide a seamless transition back to normal operation. Second,
Figure 1.2: Energy density and costs trends of commercially available Li-ion batteries for PHEV applications [2].

certain HEV or PHEV architectures allow the engine speed to be decoupled from the vehicle speed, allowing the engine to operate at an efficient speed regardless of driver demand [78]. Third, a battery pack is able to convert energy in two directions, i.e. from stored chemical energy to electrical energy or from electrical energy back to stored chemical energy. An electric machine can do the same two-directional conversion, only with electrical energy and mechanical energy. The result is that a battery/EM powertrain is able to convert the kinetic energy accumulated as the vehicle gains speed back to useful stored energy in the battery when the vehicle brakes, energy that would otherwise be lost to heat at the brake pads. Clearly this is something a liquid or gas fuel powered vehicle cannot do, as new fuel cannot be created onboard the vehicle. Finally, the tank-to-wheels efficiency of an all-electric drivetrain is significantly higher than for an ICE drivetrain (around 60% for a BEV compared to around 20% for an ICE [3]). Depending on the source of electricity, this can also result in a significantly higher well-to-wheels efficiency for a BEV. All
of these traits result in significant improvements in vehicle efficiency for hybrid or
electric vehicles that utilize battery packs over traditional ICE vehicles.

Another advantage that a powertrain with a battery pack and electric machine
brings is a boost to torque generation, particularly noticeable at low vehicle speeds
when accelerating from a standstill. Internal combustion engines are typically char-
acterized by low torque at low speeds and high torque at medium to high speeds,
whereas an electric machine is able to provide its maximum torque at low speeds
with a drop off occurring at high speeds. This makes for a natural paring in a hybrid
vehicle, in which the battery and electric machine can provide an acceleration boost
starting from rest, and the IC engine can take over at higher speeds.

Finally, an electrified vehicle can potentially have significantly lower well-to-wheels
greenhouse gas (GHG) emissions compared to a traditional gasoline or diesel powered
vehicle. For a plug-in hybrid or electric vehicle that uses electricity from the grid,
the degree to which GHG emissions can be lowered depends entirely on the source of
the electricity. For example, using electricity from the current average U.S. grid mix
(which is approximately 40% coal), an all-electric BEV can have about 20% lower
well-to-wheels GHG emissions than a gasoline HEV [54]. However, using electricity
from the grid mix in the state of Ohio (approximately 70% coal), a BEV can have
about 15% higher GHG emissions than an HEV, a somewhat counterintuitive result.

1.3 Types of Batteries

In a broad sense, batteries can be categorized as being primary or secondary. Pri-
mary batteries cannot be recharged, and must be discarded after a single discharge.
In secondary batteries, the electrochemical reactions are reversible, so they can be
recharged. Clearly secondary batteries are of interest in automotive applications, and
to date, three chemistry types have seen widespread use in the industry: lead acid,
nickel-metal hydride, and lithium-ion. Twelve volt lead acid (PbA) batteries have been widely used in traditional ICE vehicles to power the starter motor, as they are generally inexpensive, safe, and reliable [78]. However, it was not until the development of the higher power nickel-metal hydride (NiMH) battery in the late 20th century that hybrid vehicles became feasible. Although NiMH batteries have powered some of the first successful HEVs (and continue to do so), they are generally being supplanted by lithium-ion (Li-ion) batteries, which feature even higher power and energy density. As the cost of Li-ion batteries continues to fall, they look to be the dominant technology for xEVs in the near future; therefore, they are the focus of this dissertation.

1.4 Basic Function of Batteries

The basic purpose of a battery as an energy conversion device is to convert stored chemical energy into electrical energy, in the form of an electrical current. Secondary batteries like Li-ion can both supply a current to power an electrical device or use an existing current to store energy for later use. A battery consists of two electrodes, positive and negative, immersed in an electrolyte. For a Li-ion cell, the two electrodes are electrically insulated from one another by a polymer separator, which is permeable to Li-ions but not to electrons. Simply put, a Li-ion cell operates by transporting Li-ions from one electrode to the other, while simultaneously allowing electron flow through whatever external circuit is connected to the battery. Electrode materials are coated onto a current collector, a thin foil of metal that serves as a conductive pathway for electrons to travel from the electrode particles to the battery tabs. Current collectors are typically copper for the negative electrode and aluminum for the positive electrode, due to differences in each electrode potential. For most Li-ion chemistries available today, the electrolyte consists of a liquid solvent.
As the cell is discharged and charged, a reduction-oxidation (redox) reaction occurs at each electrode in which Li-ions and electrons are either the reactants or the products. A reduction reaction consumes electrons to decrease the oxidation state of a molecule, while an oxidation reaction produces electrons and increases the oxidation state [68]. During discharge, an oxidation reaction occurs at the anode, releasing an electron and a Li-ion. The anodic reaction for a typical graphite anode is:

$$Li_xC_6 \rightarrow Li^+ + C_6 + e^-$$  \hspace{1cm} (1.1)
forced to travel to the current collector and through the external circuit connected to the battery. The subscript $x$ refers to the degree of lithiation of the carbon electrode, where $0 \leq x \leq 1$. In other words, a fully lithiated carbon electrode corresponds to $x = 1$, and as the discharge continues, $x$ decreases as Li-ions are steadily removed. Simultaneously, a reduction reaction occurs at the cathode, which consumes an electron from the external circuit. The cathodic reaction for a lithium manganese oxide electrode (as an example) is:

$$Li^+ + Mn_2O_4 + e^- \rightarrow Li_yMn_2O_4$$

(1.2)

Again, the subscript $y$ refers to the degree of lithiation of the manganese oxide electrode, which increases towards $y = 1$ as the discharge progresses. For common electrode materials, the cathodic and anodic reactions represent an intercalation process, in which Li-ions are inserted into the electrode material.

A note should be made here regarding nomenclature, as the terms cathode and anode are often used to describe the positive and negative electrodes, respectively. Technically speaking, a cathode is an electrode where a reduction reaction occurs and an anode is an electrode where an oxidation reaction occurs. This means that during discharge, the positive electrode is the cathode and the negative electrode is the anode, but during charge the situation is reversed. That said, it is commonplace to use cathode to refer to the positive electrode and anode to refer to the negative electrode, and the terms will be used interchangeably in this dissertation.

1.5 Battery Electrodes

An electrode material is characterized by a specific capacity, which dictates how much charge can be stored for a given electrode mass. Specific capacity has important
implications in automotive applications, as it relates to the electric driving range that is possible for a given battery pack weight.

Electrode materials are also characterized by a standard reference potential, typically defined at 25°C, 1 atm, and electrolyte solute concentrations of 1 M. Since the concept of electric potential is only meaningful as a potential difference, the reference potential of a material must be given relative to another electrode potential. Reference potentials are often given relative to a standard hydrogen electrode (SHE). More commonly for Li-ion electrode materials, reference potentials are given relative to a pure Li metal electrode, where the pure Li metal electrode potential is defined as 0 V. So for example, the standard reference potential for a lithium-iron-phosphate electrode is 3.4 V, meaning that a cell constructed with one lithium-ion-phosphate electrode and one Li metal electrode would read 3.4 V (at standard conditions). Of course, this potential is not constant in practice, as temperature and concentrations continuously change with use.

The standard reference potentials (relative to pure Li) and specific capacities for some common Li-ion electrode materials are listed in Table 1.1.

Analyzing standard electrode potentials is useful for designing a cell, in order to select the two electrode materials. The nominal voltage across the battery tabs is the difference between each electrode potential. In general, to build a high voltage, and therefore high power cell, one would want to pair a high potential positive electrode with a low potential negative electrode. The electrode material applicable to Li-ion batteries with the lowest potential is pure Li metal itself (0 V by definition). However, pure Li is not used as a negative electrode in commercial batteries, mainly due to safety concerns. Lithium metal dendrites can grow on the pure Li electrode and could potentially break through the separator and short the cell. Instead, most commercial Li-ion batteries today use some form of graphite (0.1 V) as a negative electrode.

For a typical Li-ion cell, the electrodes consist of porous materials so that the
Table 1.1: Standard reference potential and theoretical specific capacity for different electrode materials [49,100].

<table>
<thead>
<tr>
<th>Electrode Material</th>
<th>Potential vs. Li/Li⁺ (V)</th>
<th>Specific Capacity (mAh/g)*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive Electrodes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LiMn₂O₄</td>
<td>4.1</td>
<td>100-120</td>
</tr>
<tr>
<td>LiCoO₂</td>
<td>3.9</td>
<td>140-170</td>
</tr>
<tr>
<td>LiFePO₄</td>
<td>3.45</td>
<td>170</td>
</tr>
<tr>
<td>LiNi₁/₃Mn₁/₃Co₁/₃O₂</td>
<td>3.8</td>
<td>160-170</td>
</tr>
<tr>
<td>LiNi₀.₈Co₀.₁₅Al₀.₀₅O₂</td>
<td>3.8</td>
<td>180-240</td>
</tr>
<tr>
<td>Li-Sulfur**</td>
<td>2.1</td>
<td>1280</td>
</tr>
<tr>
<td><strong>Negative Electrodes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Li₄Ti₅O₁₂</td>
<td>1.6</td>
<td>175</td>
</tr>
<tr>
<td>LiC₆ (Graphite)</td>
<td>0.1</td>
<td>372</td>
</tr>
<tr>
<td>Li-Tin**</td>
<td>0.6</td>
<td>994</td>
</tr>
<tr>
<td>Li Metal**</td>
<td>0.0</td>
<td>3862</td>
</tr>
<tr>
<td>Li-Silicon**</td>
<td>0.4</td>
<td>4200</td>
</tr>
</tbody>
</table>

*The theoretical specific capacity is listed here. For various practical reasons, the realizable capacity is often lower.

**These chemistries are not currently commercially available.

electrolyte is able to permeate into the open spaces of the electrodes. This increases the surface area of the interface between electrode particles and the electrolyte, effectively increasing the rate at which the redox reactions can occur, and therefore the power that can be extracted from the cell. However, this additional power comes at a price, as a porous electrode means there is less electrode material available to store Li in a given volume, and therefore less specific capacity. A typical electrode has a porosity of approximately 40%.

In addition to the volume lost due to the porous electrode, the electrode material is supplemented with conductive carbon particles, which act to enhance the conductive pathway between the electrode particles and the current collector, but cannot store Li themselves. A binder material is also needed to glue the electrode and carbon particles together into one solid structure. Together the carbon additive and binder
are known as inactive material, in order to differentiate with active electrode material that can store lithium. While the carbon additive and binder are necessary components, they act to reduce the effective specific capacity of an electrode by replacing potential active material with inactive material. A typical electrode is comprised of approximately 10% inactive material by volume. Therefore, when also taking into account the electrode porosity, only about 50% of the total volume of an electrode is available for storing lithium.

Composite electrodes are one particular class of electrode that are of interest in automotive applications (also referred to as blended or mixed material electrodes). Composite electrodes consist of two (or more) active materials mixed together into one electrode. This allows for more flexibility in designing a cell to meet a desired performance. For example, layered LiNi$_{1/3}$Mn$_{1/3}$Co$_{1/3}$O$_2$ (NMC), which has a high energy density but poor rate capability [66] can be mixed with LiMn$_2$O$_4$ spinel (LMO), which has great rate capability but suffers from poor life due to manganese dissolution in the electrolyte [47,56]. The result is a composite cathode that exhibits good energy and power densities with improved cycle life [70]. Essentially composite electrodes allow battery manufacturers to better balance cell energy, power, life, and cost by selecting which electrode materials to mix together, and in what ratios. However, composite electrodes offer an additional challenge from a modeling and estimation perspective. Each composite active material may not be utilized equally throughout a given drive cycle, with implications in how much power can be extracted at any given time. Furthermore, each composite active material may age at a different rate and according to a different physical mechanism, which affects how the cell capacity and power characteristics change over time. For these reasons, batteries with composite electrodes are the focus of this dissertation.
1.6 Aging Mechanisms

As batteries age over time, they tend to lose capacity and see an increase in internal resistance. In general, the loss in capacity decreases the electric range of an xEV, while a rise in internal resistance decreases the available power. Battery life can be diminished either through the accumulation of current from cycling (referred to as cycle life) or simply by the passage of time (referred to as calendar life). Depending on the cell chemistry and operating conditions, aging may be caused by different degradation mechanisms. The most relevant degradation mechanisms are described below.

- Growth of the SEI layer

Electrode particles form a surface film when in contact with the electrolyte, known as the solid-electrolyte interface (SEI) layer. The SEI layer forms as a result of a side reaction that occurs between the active electrode material and the electrolyte solvent. The products of this side reaction contain lithium, so the growth of the SEI results in the loss of cyclable lithium. Essentially, lithium that could otherwise be used to cycle back and forth between the electrodes becomes trapped in the SEI layer. The loss of cyclable lithium results in a loss in cell capacity. Another effect of SEI growth is to increase the cell internal resistance, as the SEI layer presents a barrier to lithium transport to and from the electrode surface. SEI growth occurs mainly at the negative electrode, particularly at high SOC, due to the low potential of the negative electrode, although SEI can also grow at the positive electrode [5, 95]. Additionally, SEI growth is exacerbated at high temperatures, since the rate constant of the side reaction increases with temperature (known as the Arrhenius relationship) [95].

- Active material dissolution
Active material dissolution is a prominent degradation mechanism for certain electrode active materials, particularly for Mn$_2$O$_4$ (LMO) spinel electrodes. The manganese in LMO electrodes can dissolve in the electrolyte, effectively reducing the sites available for lithium intercalation and reducing the cell capacity. It has also been shown that the dissolved Mn ions can cause further capacity loss at the negative electrode by either catalyzing SEI growth or by intercalating into the graphite, taking the place of lithium [5, 47, 50, 56, 95].

- **Binder cracking**

Another way in which active material can be lost is through stress-induced cracking of the binder that holds the electrode and conductive carbon particles together [7, 22, 25, 95]. Over time, mechanical stress from the continuous intercalation and deintercalation of lithium (resulting in slight volume expansion and compression of the electrode) or thermal stress from changes in cell temperature (also contributing to volume changes) can lead to a fracturing of the binder material. This fracturing can then lead to the degradation or even complete loss of the conductive pathway required by electrons to travel from the electrode surface to the current collector. A reduced conductive pathway effectively increases the internal resistance, and therefore decreases the power, of the cell. Once electrode particles become completely electrically isolated, they can no longer participate in the redox reactions, resulting in capacity loss.

- **Lithium plating**

Lithium plating or deposition can occur at the negative electrode when solid lithium precipitates from the liquid electrolyte onto the electrode surface [95]. This results in the loss of cyclable lithium, and therefore capacity. Lithium plating occurs if the negative electrode potential crosses below 0 V (relative to pure Li), at which point the precipitation of lithium becomes favorable.
Although the 0 V threshold is not typically reached in normal operation, even at high SOC, it can be reached during operation at low temperature.

To summarize, for the most common lithium-ion chemistries, the dominant aging mechanisms at the anode are SEI formation/growth and lithium plating, while aging at the cathode is largely due to loss of active material (either through dissolution or cracking) [95].

1.7 Battery Management System

The battery management system (BMS) in a hybrid or electric vehicle consists of both hardware and software and is in charge of all aspects of control, diagnosis, and prognosis of the battery pack. These functions include estimating the battery state-of-charge (SOC), estimating battery state-of-health (SOH), predicting miles-to-empty, determining the power demand and maximum power to or from the battery, controlling the thermal management of the pack, balancing cells within the pack, and maintaining safe operation of the cells and pack [20, 78].

The battery state of charge is a measure of the remaining useful charge in a battery, given as a percentage of its total capacity (SOC is defined formally in Chapter 4). It is analogous to the fuel gauge on a traditional ICE vehicle. Broadly speaking, a decline in state of health refers to the degradation of the battery over time, resulting in capacity and power fade (SOH is defined formally in Chapter 5). It is analogous to the size of the fuel tank on an ICE vehicle, if say the fuel tank were to shrink over time. Both SOC and SOH need to be estimated by the BMS in order to ensure safe and efficient operation of the vehicle, as well as to give the driver an accurate miles-to-empty indicator (especially critical in a BEV, where charging opportunities are limited and time consuming).
For hybrid vehicles where power can be split between the engine and the battery pack, the BMS must decide how much power to draw from the battery at any given time. The algorithm to determine the power split may be optimized for performance, drivability, and/or life, but is always dependent on knowledge of the pack SOC. For HEVs the BMS ensures that the battery operates between a narrow SOC window so as to remain charge sustaining over the long term. For PHEVs the vehicle may operate in electric-only (charge depleting) mode until the battery SOC depletes to approximately 20% or less, at which point operation continues in charge sustaining mode. The battery power demanded by the BMS must not result in over-charging or over-discharging the battery, so the maximum allowable charge and discharge rate must also be calculated and taken into account.

Battery packs must include some cooling system, which may contain active and passive elements, in order to dissipate heat that is generated while cycling the battery. Cell temperatures are monitored in order to ensure safe operation within some temperature range (typically $< 40^\circ$C) and maintain the uniformity of temperature among cells [78], although not all cell temperatures may be monitored in a pack. Battery packs may use an air or liquid cooling system, where the fluid is passed through channels in between cells. Additional passive cooling techniques may be used, such as cooling fins, to enhance heat transfer from the cells.

In general, the functions of a BMS require a model of the voltage and temperature dynamics of the cell and/or pack. A hierarchy of these battery models is described in Chapter 2. While production vehicles today use equivalent circuit models due to their relative simplicity, the focus of this dissertation is to investigate the use of physics-based models to perform SOC and SOH estimation.
1.8 Basic Terminology

Some additional terms commonly used in the battery industry and in this dissertation are defined here.

- Capacity

The capacity of a battery is the amount of charge it is capable of storing. Formally, capacity is a measure of charge (in Coulombs), but is typically given in units of Amperes x hours (Ah). The theoretical specific capacity (in Ah/kg) of an electrode material is defined as:

$$q = \frac{n_e F}{3600M}$$  \hspace{1cm} (1.3)

where $n_e$ is the valency of the material (the stoichiometric number of electrons taking part in the reaction), $F$ (C/mol) is Faraday’s constant, and $M$ (kg/mol) is the molecular weight of the active electrode material. The scaling factor, 3600, is used to convert from seconds to hours. For a number of reasons, this theoretical capacity is never achieved in practice, one of which is that a cell must operate between two voltage limits. Overcharging may lead to lithium plating at the anode or electrolyte decomposition and outgassing at the cathode, while overdischarging may lead to dissolution of the current collector at the anode or breakdown of the active material at the cathode. For these reasons, the practical capacity of a battery is dictated by the safe operating window between two voltage limits, typically between 2.0-2.8 V and 3.6-4.2 V for lithium ion batteries. The practical capacity, $C$, is obtained experimentally by integrating the applied current, $I$, to get from the upper voltage limit at time $t_0$ to the
lower limit at time $t_f$ (or vice versa):

$$C = \frac{1}{3600} \int_{t_0}^{t_f} I(t) dt \quad (1.4)$$

Typically, the practical specific capacity of an electrode (equal to $C$ normalized by the electrode mass) is on the order of 30% of the theoretical electrode specific capacity, $q$ \[^78\]. This is due to the fact that the electrode material cannot be completely lithiated or delithiated without violating the cell voltage limits. The inactive material added to the electrode structure also contributes to the lower practical specific capacity, as the inactive material adds weight while not actually providing any capacity itself.

- **C-rate**

  The C-rate refers to a specific value of current that is used to charge or discharge a battery. By definition, a C-rate of 1 (written as 1C) corresponds to the current required to completely charge or discharge a cell in 1 hour. Therefore, 2C corresponds to the current required to completely charge or discharge the cell in 0.5 hours. The amount of current required to do this depends on the cell capacity. The C-rate is related to the applied current, $I$ (in A), cell capacity, $C$ (in Ah), and time to completely charge or discharge, $T$ (in h) by: $C - rate = \frac{I}{C} = \frac{1}{T}$. The rate at which a cell is charged is or discharged is often given in terms of a C-rate instead of current in order to give a sense of the severity of the rate without needing to specify the cell capacity. For example, 100 A of current is very aggressive for a small 5 Ah cell (high C-rate of 20C), but not very aggressive for a large 50 Ah cell (only 2C).

- **Internal resistance**

  Internal resistance refers to the resistance to electron and ion flow inside a
battery that causes an initial jump in the voltage monitored across the terminals upon application of a current. For a charge current the voltage jump is positive, while for a discharge current the voltage jump is negative. The voltage jump is a consequence of Ohm’s law, $\Delta V = \Delta IR$. Formally, the internal resistance should refer to the resistance responsible for the voltage jump the instant after the current is applied; however, it is common to list a “1 second” or “10 second” resistance, which is the resistance corresponding to the voltage jump 1 s or 10 s after the current is applied.

- **Cyclable lithium versus active material**

Loss of cyclable lithium and loss of active material are two dominant degradation mechanisms for Li-ion cells, so it is important to distinguish them. Electrode active material is the host material of the intercalated lithium, such as $\text{Mn}_2\text{O}_4$ or $\text{FePO}_4$. Loss of active material results in the loss of host sites for lithium to occupy, thereby reducing the ability of the electrode to store charge. Cyclable lithium is the total number of lithium molecules that are available to participate in the redox reactions at each electrode. Consumption of lithium in a side reaction reduces the number of lithium molecules that can participate in the desired redox reaction, thereby reducing the ability of the electrodes to store charge. Both loss of active material and loss of cyclable lithium decrease the cells ability to store charge, and therefore decrease its capacity.

### 1.9 Thesis Objectives

This dissertation proposal focuses on development and validation of electrochemical model-based estimation algorithms, with application in hybrid/electric vehicles for onboard SOC and SOH estimation. The main objectives of the work are to:
1. Develop a systematic approach to designing electrochemical model-based SOH estimators for a given cell chemistry, by first determining the primary degradation mechanism(s) and then designing an appropriate estimation algorithm.

2. Understand the observability properties of the electrochemical battery model, and the consequences for observer design.

3. Experimentally validate SOC/SOH estimation algorithms applied to composite LMO-NMC batteries, using different nonlinear observers.

4. Develop methodology to quantify estimate error bounds, given uncertainty in the model and measurements.

5. Investigate the feasibility of simultaneously estimating model parameters associated with multiple degradation mechanisms.

Chapter 2 outlines a hierarchy of battery models used to relate measurable quantities like current and voltage to parameters associated with SOC and SOH, with the focus on the single particle framework used in the estimation work. Chapter 3 reviews concepts related to nonlinear state estimation, including observability and observer selection and design. Chapter 4 highlights SOC estimation results, and validation with experimental data collected from cells with a composite electrode. Chapter 5 highlights SOH estimation results, where specific model parameters related to aging mechanisms are estimated and validated with experimental data. Finally Chapter 6 summarizes the contributions in this work and suggests ways in which the analysis could be expanded in future work.
Chapter 2

Battery Modeling

2.1 Introduction

As is true for any dynamic system, being able to model the behavior of an automotive battery is useful for a number of applications, from vehicle design, to diagnostics and prognostics, to development of new battery materials. In some cases, it may be sufficient to model the current and voltage relationship without needing information about the inner workings of the battery. Computational efficiency may be prioritized at the expense of detailed information and physical representation. In other cases, it is necessary to track the interactions of individual molecules, greatly increasing the complexity of the model. For SOC and SOH estimation purposes, a battery model that includes parameters associated with these metrics is a useful tool that facilitates model-based estimation techniques. The focus of this dissertation is on electrochemical-based models that are simple enough to be used to estimate SOC and SOH onboard a vehicle, but still retain some of the physical meaning of the governing electrochemical equations.

In this chapter, a hierarchy of battery models is described, with a focus on the porous electrode model and its various reduced-order forms. Finally, motivation is
provided for using these electrochemical models for SOC and SOH estimation. Much of this chapter summarizes the work done by Jim Marcicki in [58], which provides the modeling foundation for the estimation work done in this dissertation.

2.2 A Hierarchy of Battery Models

There are different categories of battery models of various complexities. Figure 2.1 shows a hierarchy of models that describe a battery at length scales ranging from the atomic level to a complete cell. Generally speaking, the smaller the length scale, the larger the computational effort required to simulate the model.

Computational methods attempt to simulate the interactions between individual atoms. Rather than considering properties at the macroscopic level like temperature or pressure, molecular dynamics (MD) attempts to model the behavior of individual atoms at the microscopic level. Naturally, this is a very computationally intensive
process, so the physical space that is simulated must be carefully constrained. Modeling molecular interactions involves using Newton’s laws of motion to solve for the positions of atoms and their momenta, and defining a microscopic state that incorporates these quantities for all \( N \) atoms involved in the simulation. Relating this microscopic state back to a macroscopic state, defined by say temperature and pressure within a given volume, can be difficult as there are many microscopic states that correspond to a single macroscopic state. In fact, macroscopic states represent the ensemble average of all possible sets of corresponding microscopic states. MD simulations can be used to characterize ion transport mechanisms [46] or understand how side reactions form the SEI layer [10].

Density functional theory (DFT) uses quantum mechanics to approximate the solution to the many-particle Schrödinger equation, i.e. obtain their quantum state, by calculating the electronic structure (electron density) of the particles [62]. DFT can be used to predict the performance of new electrode materials, including both the nominal characteristics and how they might change with age and usage.

Porous electrode models, and their various simplifications, are covered in detail in the following sections. In short, porous electrode models consider a macroscopic view of transport and thermodynamics that govern the current and voltage behavior of a cell. These models make certain assumptions about the electroneutrality of species and the structure of the electrode particles, but still retain the physical meaning of model parameters. The single particle model is a common approximation of porous electrode theory, that considers a single electrode particle, representing the average behavior of the entire electrode.

Equivalent circuit models are an example of an analogy-based approach to modeling. Rather than consider the physical processes occurring in the battery, a circuit model attempts to mimic the battery input-output behavior by constructing a model of a hypothetical electrical circuit that has a similar input-output behavior. This
is a common technique when modeling mechanical, fluid, or thermal systems for the purposes of control, as the concepts of resistance, capacitance, and inductance are often more familiar to controls-oriented engineers. A basic equivalent circuit battery model consists of a voltage source (representing the cell OCV), a series resistance (representing the Ohmic resistance losses in the electrode and electrolyte), and a resistor-capacitor pair in parallel (representing the relaxation phenomena). The values of the voltage source, resistance, and capacitance are often scheduled through a lookup table based on SOC, current, temperature, and/or current direction [19, 20, 24, 36, 38, 42–44, 55, 71–73, 80]. Other circuit elements may be added to improve the model accuracy, such as additional resistor-capacitor pairs or a Warburg element to better represent the diffusion dynamics. This type of model is not completely without physical meaning in battery applications, as both the circuit model and the actual battery involve concepts like voltage, current, and resistance; however, the dynamics of mass transport and reaction kinetics lose their physical interpretation in a circuit model.

Clearly there is a trade-off between computational effort and the accuracy and/or physical representation of the model. For the purposes of online state estimation, only equivalent circuit models and single particle models may be practical, as the computing power of the BMS hardware is limited.

2.3 Nomenclature

The following variables are used in this chapter to develop the porous electrode model.

- $A$ Current collector area (m$^2$)
- $\bar{D}_{ij}$ Diffusion coefficient describing interaction between species $i$ and $j$ ()
- $D_e$ Liquid diffusion coefficient based on a gradient in concentration (m$^2$/s)
- $\bar{D}_e$ Electrolyte diffusion coefficient based on a gradient in
chemical potential ()

$D_i$ Solid diffusion coefficient (m²/s)

$E$ Electric field (V/m)

$F$ Faraday’s constant (C/mol)

$G$ Gibbs free energy (J)

$I$ Current (A)

$K$ Species friction coefficient (Js/m⁵)

$L_{cell}$ Total cell thickness (m)

$L_i$ Electrode thickness (m)

$L_{sep}$ Separator thickness (m)

$N$ Flux density of species $i$ (mol/m²/s)

$\bar{R}$ Universal gas constant (J/K/mol)

$R_c$ Ohmic resistance (Ω)

$R$ Particle radius (m)

$T$ Cell temperature (K)

$U$ Electrode open circuit voltage (V)

$V$ Cell voltage (V)

$V_e$ Liquid phase voltage across the cell (V)

$a$ Electrode specific interfacial area (m²)

$act$ Activity of participating species ()

$c$ Lithium concentration (mol/m³)

$\bar{c}$ Bulk particle lithium concentration (mol/m³)

$c_T$ Sum of all species concentrations (mol/m³)

$c_e$ Lithium electrolyte concentration (mol/m³)

$c_{max}$ Lithium electrode saturation concentration (mol/m³)

$c_s$ Lithium concentration at the particle surface (mol/m³)

$i_0$ Exchange current density (A/m²)

$i_e$ Liquid phase current density (A/m²)

$i_s$ Solid phase current density (A/m²)

$j$ Transfer current density (A/m²)

$k$ Reaction rate constant (m².5/mol⁰.5/s)

$n_e$ Stoichiometric number of electrons transferred ()

$n_i$ Number of moles of species $i$ ()

$r$ Coordinate along the particle radius, originating at the particle center

$r_i$ Reaction rate of redox reaction $i$ (mol/m²/s)

$s_i$ Stoichiometric coefficient of $i$ for a generic reaction

$t$ Time (s)

$i_0^i$ Lithium-ion transference number of species $i$ ()

$u_i$ Ion mobility of species $i$ (m²mol/J/s)

$v$ Bulk fluid velocity (m/s)

$v_i$ Species velocity (m/s)

$x$ Coordinate through-the-thickness of the cell,
originating at the anode current collector

\( z \) Charge number of species \( i \) ()
\( \alpha \) Reaction symmetry factor ()
\( \beta_i \) Current split factor ()
\( \gamma \) Activity coefficient in electrolyte ()
\( \epsilon_e \) Liquid phase volume fraction ()
\( \epsilon_i \) Electrode active material volume split ()
\( \eta \) Kinetic overpotential (V)
\( \mu \) Chemical potential (J/mol)
\( \bar{\mu} \) Electrochemical potential (J/mol)
\( \nu \) Number of ions produced by the dissociation of one molecule of salt ()
\( \sigma_e \) Electrolyte conductivity (S/m)
\( \phi_e \) Electrolyte potential (V)
\( \phi_i \) Electrode potential (V)

Subscript \( i \) Refers to electrode p, n, LMO, or NMC
Subscript 0 Indicates standard state
Subscript \( R \) Indicates reduced species
Subscript \( O \) Indicates oxidized species
Subscript + Indicates cation
Subscript – Indicates anion
Subscript 0 Indicates solvent

2.4 The Porous Electrode Model

The goal of any electrochemical model in the context of SOC/SOH estimation is to relate physical concepts like Li concentrations and active electrode material to the measurable inputs and outputs of a battery, namely current, voltage, and temperature. The porous electrode model [30, 68, 69], applicable to many common Li-ion battery chemistries that have porous electrodes with electrolyte filling the gaps, is able to do this by taking a macroscopic approach. The actual complex geometry of the electrode pores is disregarded; instead, the electrode is characterized by some specific interfacial area, \( a \), which is the surface area of the pores per unit volume, representative of the whole electrode. The porous electrode model also assumes electroneutrality in both solid and liquid phases, in that within a given volume element, the sum of charged species is zero. This is a valid assumption at macroscopic length
scales over which the porous electrode model is typically discretized and solved, but does not necessarily apply at the nanoscale. For example, electroneutrality does not hold near a thin (a few nanometers) region at the boundary between the electrode and electrolyte, known as the electric double layer [58]. A consequence of the assumption of electroneutrality is that the total current density is always constant, i.e. the gradient of the total current density is equal to 0. Current in this case refers to the flow of any charged particle (positive or negative). The charge carrier in the solid electrode is the electron as it travels to or from the current collectors, while the charge carrier in the liquid electrolyte is the lithium ion, Li$^+$, as it travels back and forth between each electrode. So in the porous electrode, where both solid and liquid phase are present, the sum of each form of current is constant, i.e.:

$$\nabla \cdot \vec{i}_s + \nabla \cdot \vec{i}_e = 0$$

(2.1)

where the current densities (A/m$^2$) are in general 3 dimensional vectors in space. Note that at the current collector $\vec{i}_e = 0$ and only electrons flow, since the electrolyte does not penetrate the solid current collector. Alternatively, the separator is electrically insulating, so in this region $\vec{i}_s = 0$ and only Li-ions flow. When a cell is cycled, current is transferred between the solid and liquid phase via the redox reactions at each electrode, represented by a transfer current density, $j$ (A/m$^2$), defined by:

$$a\vec{j} = \nabla \cdot \vec{i}_s = -\nabla \cdot \vec{i}_e$$

(2.2)

where $a$ is the specific interfacial area (1/m). The transfer current density describes how current is transferred from acting in one phase to acting in the other, with the understanding that any transfer of current into the solid phase has to be coupled with an equal and opposite reduction in current in the liquid phase.
2.4.1 1+1D porous electrode model

A common assumption when implementing the porous electrode model is to first assume that current density, and therefore potential, gradients only exist in the direction through-the-thickness of the cell, referred to as the $x$ direction (see Fig. 2.2). Then the transfer current density, $j(x,t)$, becomes only a function of $x$ and time. The solid electrode particles can also be approximated as spheres oriented along the $x$ direction, making the specific interfacial area in electrode $i$ equal to:

$$a_i = \frac{3\epsilon_i}{R_i}.$$  \hspace{1cm} (2.3)

where $R_i$ (m) is the particle radius and $\epsilon_i$ is the active material volume fraction. By discretizing the $x$ direction into discrete nodes, a spherical particle can be assigned to each node, subject to a transfer current density evaluated at the specific $x$ coordinate. If the spherical particles are assumed to have gradients only in the radial direction, $r$, the final result is a model that spans two spatial dimensions: $x$ and $r$ (the so-called $1+1$D model). The $x$ coordinate starts from $x = 0$ at the negative electrode current collector and ends at $x = L_{\text{cell}}$ at the positive electrode current collector,
where \( L_{\text{cell}} = L_n + L_{\text{sep}} + L_p \) is the sum of negative electrode, separator, and positive electrode thicknesses.

### 2.4.2 Governing electrochemical equations

Using the 1+1D framework, modeling the unit cell dynamics requires consideration of the thermodynamics, kinetics of electrochemical reactions, and mass transport occurring in the cell. Thermodynamics is used to describe an electrode’s open circuit voltage (OCV), which is the potential difference between two electrodes if the cell is at equilibrium with no current flowing. In general, the OCV is a function of the cell SOC and temperature. Then, kinetics and mass transport describe how the behavior of the cell changes once current begins to flow. Electrochemical reactions and diffusion of ions can only happen so fast, and resistance to electron and ion flow inherently generates heat. In essence, the OCV represents the ideal voltage that is possible for given conditions, while kinetics and transport take into account the deviation from that ideal voltage once current is introduced.

### 2.4.3 The open circuit voltage from thermodynamics

First, as was mentioned in Chapter 1, it is important to note that the absolute value of potential is not measurable without specifying a reference point. The same thing is true for other concepts like gravitational potential energy, equal to \( mg\Delta h \), where \( \Delta h \) is the height difference between two points. So in order to talk about the potential of a given electrode, the other electrode that would be needed to form a complete cell must be specified. The OCV of a Li-ion electrode is typically given with respect to a pure Li metal electrode reference. For example, the OCV of a lithium iron phosphate
electrode would be defined as:

$$OCV_{\text{LiFePO}_4} = \Phi_{\text{LiFePO}_4} - \Phi_{\text{Li}}$$  \hspace{1cm} (2.4)$$

where the individual values of potential, $\Phi_{\text{LiFePO}_4}$ and $\Phi_{\text{Li}}$, are not measurable, but their measured difference gives the OCV of the electrode pair. The term potential is used to characterize an individual electrode or a certain location within an electrode or electrolyte, while the term voltage refers to a potential difference between two electrodes, which is actually measurable. As a convention, the potential of $\Phi_{\text{Li}}$ can be defined as 0 V (in the same way that the ground in an electrical circuit may be defined as 0 V). In this way, the OCV of the LiFePO$_4$ electrode can be listed as being, say 3.4 V, with the understanding that this is really the potential difference that would be measured between a cell made with LiFePO$_4$ and pure Li electrodes.

Second, while the OCV may be a property of a single electrode of interest paired with a pure Li reference, a full cell is made up of two electrodes, for example LiFePO$_4$ and graphite, each with their own OCVs with respect to pure Li. The resting voltage that is measured between the cell terminals is equal to the difference between each electrode OCV. In this case, the potential of pure lithium cancels out:

$$OCV_{\text{cell}} = (\Phi_{\text{LiFePO}_4} - \Phi_{\text{Li}}) - (\Phi_{\text{LiC}_6} - \Phi_{\text{Li}}) = \Phi_{\text{LiFePO}_4} - \Phi_{\text{LiC}_6}$$  \hspace{1cm} (2.5)$$

At a systems level, the term OCV is often assumed to mean the complete cell OCV, incorporating the combined effect of both electrode OCVs.

Deriving the OCV of an electrode in the context of thermodynamics requires starting from the relationship between a change in Gibbs free energy, $\Delta G$ (J), and
the equilibrium or open circuit potential, $U$ (V) [68, 78]:

$$
\Delta G = -n_e F U
\quad (2.6)
$$

where $n_e$ is the stoichiometric number of electrons transferred in the reaction and $F$ (C/mol) is Faradays number. The Gibbs free energy is the useful work that can be obtained from a thermodynamic system (an electrochemical reaction in this case). In general, chemical species will always try to minimize the free energy of a system; therefore, if the $\Delta G$ between reactants and products is negative, the reaction will proceed spontaneously in order to reach a lower energy state. Conversely if the $\Delta G$ is positive, then the reaction is not favored, and can only proceed if energy is added to overcome this barrier. If the $\Delta G = 0$, then the reaction is neither favored nor disfavored, and the system is at equilibrium. For example, in an electrochemical cell, the $\Delta G$ between reactants and products for the oxidation and reduction reactions at the negative and positive electrodes during a discharge is negative, so the reactions proceed spontaneously. However, the $\Delta G$ of the reverse reactions during a charge is positive, indicating that the reaction can only be driven by applying an external potential between the electrodes.

A superscript 0 is used to indicate the reaction is at standard state (electrolyte solutes are at 1 M, $p = 1$ bar, $T = 25^\circ C$):

$$
\Delta G^0 = -n_e F U^0
\quad (2.7)
$$

If the reaction occurs at non-standard conditions, an additional term is needed to calculate $\Delta G$, related to the temperature and activity, $act$, of the reactants and products:

$$
\Delta G = \Delta G^0 + \bar{R}T \ln \left( \frac{\prod act_{prod}}{\prod act_{react}} \right)
\quad (2.8)
$$
where $\bar{R}$ is the universal gas constant (J/K/mol). Activity is a measure of the effective concentration, after accounting for molecular interactions. Formally, activity is defined as: 

$$a_{ci} = \exp \left( \frac{\mu_i - \mu_0}{\bar{R}T} \right),$$

where $\mu_i$ (J/mol) is the chemical potential of species $i$ (subscript 0 indicating standard state). Chemical potential is discussed further in Section 2.4.7, but it is related to the concentration of species. So the variation in OCV as a function of the concentration of intercalated lithium in an electrode is due to the variation in the chemical potential, and therefore activity, of reactants and/or products as these concentrations change. Finally, the potential at a non-standard state can be written by substituting Eq. (2.6) and (2.7) into Eq. (2.8):

$$\Delta G = \Delta G^0 + \bar{R}T \ln \left( \frac{\prod a_{prod}}{\prod a_{react}} \right) = -n_e F U$$

(2.9)

$$-n_e F U^0 + \bar{R}T \ln \left( \frac{\prod a_{prod}}{\prod a_{react}} \right) = -n_e F U$$

(2.10)

$$U = U^0 - \frac{\bar{R}T}{n_e F} \ln \left( \frac{\prod a_{prod}}{\prod a_{react}} \right)$$

(2.11)

This is called the Nernst equation. Note again that this is not a potential that can be measured. Only the potential difference between two electrodes (say the electrode of interest coupled with a pure lithium reference electrode) is measurable. For this reason, the open circuit potential of an electrode is given instead as an open circuit voltage between some reference electrode (for example in Table 1.1 given versus a pure Li reference). The standard open circuit potential, $U^0$, is constant and is a property of the specific electrode material (Table 1.1), but the actual potential deviates from this constant depending on the temperature and species activity. Theoretically, Eq. (2.11) can be used to calculate an electrode pair OCV as a function of lithium concentration and temperature. However, calculation of the activity of reaction species can be difficult, and often requires a model based on empirical parameters [94]. Instead,
for battery modeling purposes the OCV dependence on lithium concentration and
temperature is usually defined experimentally by constructing a half cell with the
electrode of interest versus pure Li. Cycling the half cell at a very low rate (<
C/20) gives a good approximation of the OCV across a range of concentrations,
since the resistive losses are small at such low rates. For example, Fig. (2.3) shows an
experimentally obtained OCV curve of a graphite half cell versus Li content, cycled at
C/20. Multiple OCV curves can be constructed by testing at different temperatures.
An alternative method for experimentally obtaining the OCV is to charge or discharge
the half cell in multiple pulses, allowing the cell to rest and come to equilibrium in
between each pulse, and recording the resting voltage at each step. This method can
give a more accurate measure of the true OCV at a given resting point since the cell
is allowed to come to equilibrium prior to recording the potential; however, the test
may take a much longer time to conduct, depending on how many data points per
curve are desired.

After establishing the electrode open circuit voltage when no net current is flowing,
the next step is to quantify the various sources of resistive losses, or overpotentials,
when current is introduced. These include ohmic, kinetic (or activation), and concen-
tration overpotentials, and can be categorized as acting in the solid electrode, liquid
electrolyte, or at the interface between the two.

2.4.4 Kinetic overpotential

The kinetics of a reaction take into account the fact that a reaction (i.e. obtaining
products from reactants) can only happen at a certain rate, dependent on the react-
ing species and environmental conditions. Reactions are limited by their ability to
overcome an energy barrier (known as the activation energy) in order to reach an
overall lower energy state. Figure 2.4 shows a schematic of this process, in terms of
At each electrode, both anodic (oxidizing) and cathodic (reducing) reactions occur continuously; however, one direction is always favored if the cell is not at equilibrium. The reaction rate of a generic reaction, \( r = kc \), depends on the rate constant, \( k \) (m/s), and the concentration of reduced or oxidized species, \( c_{s,R/O} \) (mol/m\(^3\)), which in the case of an electrochemical reaction, is evaluated at the electrode surface. The Arrhenius equation gives the relationship between the reaction rate constant and the activation energy, \( \Delta G_{\text{act}} \) (J), for both anodic and cathodic reactions [68], resulting in the following expressions for reaction rates:

\[
r_a = c_{s,R} k_a = c_{s,R} k_{0,a} \exp \left( -\frac{\Delta G_{\text{act},a}}{RT} \right) \tag{2.12}
\]

\[
r_c = c_{s,O} k_c = c_{s,O} k_{0,c} \exp \left( \frac{\Delta G_{\text{act},c}}{RT} \right) \tag{2.13}
\]

Figure 2.3: Open circuit voltage versus Li concentration of a graphite electrode.
Figure 2.4: Schematic of the change in Gibbs free energy going from reactants to products as a function of the reaction coordinate.

where $k_0$ (m/s) is the standard reaction rate constant and subscript $a$ and $c$ indicate the anodic and cathodic reactions, respectively. If more than one reactant is present, $c_{s,R/O}$ is the product of all reactant concentrations. The exponential term represents the probability that the energy barrier can be overcome by a given set of reactants, in order to form products. Note that a higher concentration of reactants, a higher temperature, and a lower activation energy all act to increase the reaction rate. Substituting Eq. (2.6) for the definition of a change in Gibbs free energy gives:

$$r_a = c_{s,R}k_{0,a} \exp \left( \frac{n_e F(1-\alpha)\Delta U}{RT} \right)$$ \hspace{1cm} (2.14)

$$r_c = c_{s,O}k_{0,c} \exp \left( -\frac{n_e F\alpha \Delta U}{RT} \right)$$ \hspace{1cm} (2.15)

where $\Delta U = U - U^0$ is the change in electrode potential from the standard equilibrium potential, as defined previously. An additional factor, $\alpha$, is added as a measure of the symmetry of the energy barrier between anodic and cathodic reactions (typically
\( \alpha = \frac{1}{2} \) is a good approximation). Since both anodic and cathodic reactions occur simultaneously, the net reaction rate is:

\[
r = r_a - r_c = c_{s,R}k_{0,a} \exp \left( \frac{n_e F(1 - \alpha)\Delta U}{RT} \right) - c_{s,O}k_{0,c} \exp \left( -\frac{n_e F\alpha \Delta U}{RT} \right)
\]

(2.16)

As a convention, Eq. (2.16) would be true at the positive electrode during discharge, while at the negative electrode the sign of the total rate would be reversed, i.e. \( r = r_c - r_a \). Finally, the reaction rate can be written in terms of a current density, \( j \) (A/m\(^2\)), according to \( j = n_e F r \), and the final expression for the net current flow becomes:

\[
\dot{j} = n_e F \left( c_{s,R}k_{0,a} \exp \left( \frac{n_e F(1 - \alpha)\Delta U}{RT} \right) - c_{s,O}k_{0,c} \exp \left( -\frac{n_e F\alpha \Delta U}{RT} \right) \right)
\]

(2.17)

This is the well-known Butler-Volmer equation [68,78].

For the Li-ion cell, the reactant concentrations can be defined specifically for the following redox reaction:

\[
Li_xC_6 \rightarrow Li^+ + C_6 + e^- \quad (2.18)
\]

\[
Li^+ + M + e^- \rightarrow Li_yM \quad (2.19)
\]

where the negative electrode is assumed to be graphite and the positive electrode is comprised of a generic metal or metal oxide represented by \( M \). Equation (2.17) can be replicated for both positive and negative electrodes:

\[
\dot{j}_p = Fk_{0,p} \left( c_{s,R,p} \exp \left( \frac{F\Delta U_p}{2RT} \right) - c_{s,O,p} \exp \left( -\frac{F\Delta U_p}{2RT} \right) \right)
\]

(2.20)

\[
\dot{j}_n = Fk_{0,n} \left( c_{s,O,n} \exp \left( -\frac{F\Delta U_n}{2RT} \right) - c_{s,R,n} \exp \left( \frac{F\Delta U_n}{2RT} \right) \right)
\]

(2.21)

with the assumption that \( \alpha = \frac{1}{2} \), \( n_e = 1 \) for the lithium ion reaction, and the rate
constants for anodic and cathodic reactions at a given electrode are equal, to simplify the derivation.

When defining concentrations of each set of reactants, it is useful for the model to frame them in terms of a saturation concentration of the lithiated active material, $c_{\text{max},i}$. This is simply a change in variables to help simplify the notation. In other words define:

\[
c_{\text{max},\text{Li}C_6} = c_{C_6} + c_{\text{Li}xC_6}
\]

\[
c_{\text{max},\text{Li}M} = c_{M} + c_{\text{Li}yM}
\]

The maximum concentration of LiC$_6$ or LiM (when it is saturated with Li) is equal to the sum of the delithiated active material and the lithiated active material, at any given time. Using this definition, the concentration terms can be defined as the product of all oxidized and reduced reactants for each electrode:

\[
c_{s,O,p} = c_{\text{Li}yM}
\]

\[
c_{s,R,p} = c_{\text{Li}C_6} = c_{\text{Li}y}^+(c_{\text{max},\text{Li}M} - c_{\text{Li}yM})
\]

\[
c_{s,O,n} = c_{\text{Li}xC_6}
\]

\[
c_{s,R,n} = c_{\text{Li}C_6} = c_{\text{Li}x}(c_{\text{max},\text{Li}C_6} - c_{\text{Li}xC_6})
\]

It is also important to note that these are local concentrations present at the reaction sites, namely at the electrode surface. To account for this, the additional definitions are introduced:

\[
c_{\text{Li}yM} = c_{s,p}
\]

\[
c_{\text{max},\text{Li}M} = c_{\text{max},p}
\]

\[
c_{\text{Li}xC_6} = c_{s,n}
\]
\[ c_{\text{max,LiC}_6} = c_{\text{max,n}} \] (2.31)

where subscript \( s \) indicates the concentration at the electrode surface, and subscripts \( p \) and \( n \) indicate positive and negative electrodes, respectively. The local concentration of Li\(^+\) ions in the electrolyte near each electrode surface is redefined as either \( c_{\text{Li}^+} = c_{e,p} \) or \( c_{\text{Li}^+} = c_{e,n} \) depending on the electrode in question. Substituting these redefined concentrations gives:

\[
j_p = F k_{0,p} \left( c_{e,p}(c_{\text{max},p} - c_{s,p}) \exp \left( \frac{F \Delta U_p}{2RT} \right) - c_{s,p} \exp \left( -\frac{F \Delta U_p}{2RT} \right) \right) \quad (2.32)
\]

\[
j_n = F k_{0,n} \left( c_{e,n}(c_{\text{max},n} - c_{s,n}) \exp \left( -\frac{F \Delta U_n}{2RT} \right) - c_{s,n} \exp \left( \frac{F \Delta U_n}{2RT} \right) \right) \quad (2.33)
\]

As a final step, rather than express the Butler-Volmer equation in terms of \( \Delta U \), where the reference is the standard equilibrium potential, it is often more useful to express it in terms of the equilibrium OCV, which can be obtained experimentally or from Eq. (2.11). At equilibrium, the anodic and cathodic reaction rates become equal and the net current \( j = 0 \) for each electrode. This gives:

\[
c_{e,p}(c_{\text{max},p} - c_{s,p}) \exp \left( \frac{F \Delta U_p}{2RT} \right) = c_{s,p} \exp \left( -\frac{F \Delta U_p}{2RT} \right) \quad (2.34)
\]

\[
c_{s,n} \exp \left( \frac{F \Delta U_n}{2RT} \right) = c_{e,n}(c_{\text{max},n} - c_{s,n}) \exp \left( -\frac{F \Delta U_n}{2RT} \right) \quad (2.35)
\]

Solving for \( \Delta U_p \) and \( \Delta U_n \) gives the equilibrium potentials:

\[
\Delta U_{p,eq} = \frac{\bar{R}T}{F} \ln \left( \frac{c_{s,p}}{c_{e,p}(c_{\text{max},p} - c_{s,p})} \right) \quad (2.36)
\]

\[
\Delta U_{n,eq} = \frac{\bar{R}T}{F} \ln \left( \frac{c_{e,n}(c_{\text{max},n} - c_{s,n})}{c_{s,n}} \right) \quad (2.37)
\]

The activation or charge transfer overpotential for each electrode is defined as a
change in potential from equilibrium:

\[ \eta_p = \Delta U_p - \Delta U_{p,eq} \quad (2.38) \]

\[ \eta_n = \Delta U_n - \Delta U_{n,eq} \quad (2.39) \]

Substituting for \( \Delta U_p \) and \( \Delta U_n \) in terms of the overpotential and equilibrium potential simplifies to:

\[ j_p = Fk_0,pc_e,p(c_{\text{max},p} - c_{s,p}) \exp \left( \frac{F\eta_p}{2RT} + \frac{1}{2} \ln \left( \frac{c_{s,p}}{c_e,p(c_{\text{max},p} - c_{s,p})} \right) \right) \]

\[ - Fk_0,pc_{s,p} \exp \left( - \frac{F\eta_p}{2RT} - \frac{1}{2} \ln \left( \frac{c_{s,p}}{c_e,p(c_{\text{max},p} - c_{s,p})} \right) \right) \quad (2.40) \]

\[ j_n = Fk_0,nc_e,n(c_{\text{max},n}c_{s,n}) \exp \left( - \frac{F\eta_n}{2RT} - \frac{1}{2} \ln \left( \frac{c_{e,n}(c_{\text{max},n} - c_{s,n})}{c_{s,n}} \right) \right) \]

\[ - Fk_0,nc_{s,n} \exp \left( \frac{F\eta_n}{2RT} + \frac{1}{2} \ln \left( \frac{c_{e,n}(c_{\text{max},n} - c_{s,n})}{c_{s,n}} \right) \right) \quad (2.41) \]

Removing the natural log from the exponential function gives:

\[ j_p = Fk_0,pc_e,p(c_{\text{max},p} - c_{s,p}) \exp \left( \frac{F\eta_p}{2RT} \right) \left( \frac{c_{s,p}}{c_e,p(c_{\text{max},p} - c_{s,p})} \right)^{\frac{1}{2}} \]

\[ - Fk_0,pc_{s,p} \exp \left( - \frac{F\eta_p}{2RT} \right) \left( \frac{c_{s,p}}{c_e,p(c_{\text{max},p} - c_{s,p})} \right)^{-\frac{1}{2}} \quad (2.42) \]

\[ j_n = Fk_0,nc_e,n(c_{\text{max},n}c_{s,n}) \exp \left( - \frac{F\eta_n}{2RT} \right) \left( \frac{c_{e,n}(c_{\text{max},n} - c_{s,n})}{c_{s,n}} \right)^{-\frac{1}{2}} \]

\[ - Fk_0,nc_{s,n} \exp \left( \frac{F\eta_n}{2RT} \right) \left( \frac{c_{e,n}(c_{\text{max},n} - c_{s,n})}{c_{s,n}} \right)^{\frac{1}{2}} \quad (2.43) \]

Combining terms gives:

\[ j_p = Fk_0,pc_{s,p}c_e,p(c_{\text{max},p} - c_{s,p})^{\frac{1}{2}} \left( \exp \left( \frac{F\eta_p}{2RT} \right) - \exp \left( - \frac{F\eta_p}{2RT} \right) \right) \quad (2.44) \]
Reaction occurs at the surface

Electrolyte

Electrode particle

Figure 2.5: Schematic of the concentration gradient that forms as the redox reactions occur at the electrode particle surface.

\[
j_n = Fk_{0,n} (c_{s,n} c_{e,n} (c_{\text{max},n} - c_{s,n}))^{\frac{1}{2}} \left( \exp \left( -\frac{F\eta_n}{2RT} \right) - \exp \left( \frac{F\eta_n}{2RT} \right) \right) \tag{2.45}
\]

Finally using the definition of a hyperbolic sine and defining an exchange current density, \( i_0 \), the final expressions become:

\[
j_p = 2i_{0,p} \sinh \left( \frac{F\eta_p}{2RT} \right) \tag{2.46}
\]

\[
j_n = -2i_{0,n} \sinh \left( \frac{F\eta_n}{2RT} \right) \tag{2.47}
\]

where \( i_{0,i} = Fk_{0,i} (c_{s,i} c_{e,i} (c_{\text{max},i} - c_{s,i}))^{\frac{3}{2}} \).

2.4.5 Mass transport in the spherical electrode particle and the concentration overpotential

In the porous electrode model, the electrode particles are represented as a collection of solid spheres arranged along the \( x \) coordinate of the cell. The redox reactions are assumed to occur at the surface of these spheres where there is direct contact with the electrolyte, as shown in Fig. 2.5. As lithium reacts at the surface and begins to intercalate or deintercalate to/from the particle, it causes either an accumulation or a
depletion of lithium near the surface, thereby creating a concentration gradient in the particle. In order for the reactions to continue, lithium must diffuse from the center or bulk of the particle up to the surface. In this way, the instantaneous power that can be extracted from the battery is effectively limited by the time it takes for lithium to reach the reaction sites. The concentration gradient that forms in the particle can be modeled by Fick’s law of diffusion in spherical coordinates, with the assumption that the gradient is only in the radial direction [78]:

\[ \frac{\partial c_i}{\partial t} = \frac{D_i}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial c_i}{\partial r} \right) \]  (2.48)

where \( c_i \) (mol/m\(^3\)) is the lithium concentration for particle \( i \), \( t \) (s) is time, \( D_i \) (m\(^2\)/s) is the diffusion coefficient of lithium in the solid (assumed to be constant), and \( r \) is the radial coordinate through the sphere, spanning from 0 to the particle radius, \( R_i \) (m). Fick’s law is a consequence of mass conservation within a control volume, in this case the spherical particle. The boundary conditions are defined to maintain symmetry about the particle center and impose a flux at the particle surface related to the current density at the given \( x \) coordinate:

\[ \left. \frac{\partial c_i}{\partial r} \right|_{r=0} = 0, \quad D_i \left. \frac{\partial c_i}{\partial r} \right|_{r=R_i} = \mp \frac{j_i(x,t)}{F} \]  (2.49)

where the sign of the current density is negative for a negative electrode particle and positive for a positive electrode particle. The concentration gradient in the solid has two primary effects on the cell voltage dynamics. It affects both the charge transfer overpotential and the evaluation of the open circuit voltage, since these depend on the surface concentration. The latter is known as a concentration overpotential. As the surface concentration deviates from the bulk concentration upon application of current, the OCV, evaluated at the surface concentration, deviates as well. The
change in OCV going from evaluating at the bulk concentration to evaluating at the surface concentration represents the concentration overpotential. The slope of the OCV at a given point dictates how much overpotential results from a given concentration gradient.

### 2.4.6 Electrical conduction in the solid

The flow of electrons in the solid phase is governed by Ohm’s law:

\[
\sigma \frac{\partial^2 \Phi}{\partial x^2} = a_i j_i(x, t) \tag{2.50}
\]

where \( \Phi \) (V) is the solid phase potential and \( \sigma \) (Ω/m) is the effective solid conductivity, assumed to be independent of the \( x \) coordinate. The \( x \) coordinate spans the cell from the negative electrode current collector at \( x = 0 \) to the positive electrode current collector at \( x = L_{cell} \), as shown in Fig. 2.2. Equation (2.50) is applied separately to each electrode, with boundary conditions at the current collectors and separator:

\[
-\sigma_n \frac{\partial \Phi_n}{\partial x} \bigg|_{x=0} = \sigma_p \frac{\partial \Phi_p}{\partial x} \bigg|_{x=L_{cell}} = \frac{I(t)}{A} \tag{2.51}
\]

where \( I \) (A) is the applied current and \( A \) (m²) is the current collector area.

### 2.4.7 Lithium transport in the electrolyte

Similar to in the solid phase, lithium must also be transported through the liquid electrolyte from the surface of one electrode to the other, in order to maintain current flow. Lithium can be transported in the liquid by diffusion, migration, and convection. Lithium ions entering and exiting the electrolyte create a concentration gradient in the \( x \) direction. In general, there is an excess of lithium ions near the electrode where the lithium is being oxidized (entering the electrolyte), and a shortage of lithium
ions near the electrode where the lithium is being reduced (exiting the electrolyte). The resulting concentration gradient induces diffusion of lithium from areas of high concentration to areas of low concentration. The presence of an electric field, i.e. a potential difference across the electrolyte, causes lithium ions (positively charged) to migrate towards the more negatively charged electrode, against the gradient in potential. This generally helps to transport lithium ions towards the cathode (the electrode where electrons are being reduced). Finally, convection can transport ions through bulk fluid motion; however, its effect is usually negligible in Li-ion batteries.

For a dilute electrolytic solution, the flux density of species $i$, $N_i$ (mol/m²/s), has contributions from diffusion (1st term), migration (2nd term), and convection (3rd term):

$$N_i = -D_i \frac{\partial c_i}{\partial x} - z_i u_i F c_i \frac{\partial \Phi_e}{\partial x} + c_i v$$ (2.52)

where $c_i$ (mol/m³) is the concentration of species $i$ in the liquid, $D_i$ (m²/s) is the diffusion coefficient, $z_i$ is the charge number of the ion, $u_i$ (m² mol/J/s) is the ion mobility, $\Phi_e$ (V) is the electrostatic potential (whose gradient is equal to the negative of the electric field, i.e. $\frac{\partial \Phi_e}{\partial x} = -E$), and $v$ (m/s) is the bulk fluid velocity. For a dilute solution, the ionic mobility of species $i$ is given by the Nernst-Einstein equation: $u_i = \frac{D_i \bar{R} T}{F}$. The assumption here is that there are no significant interactions among the various species in the dilute solution, only between each species and the solvent.

Conservation of mass dictates how this flux changes the lithium concentration over time:

$$\epsilon_e \frac{\partial c_e}{\partial t} = a_i j_i(x, t) - \frac{\partial N}{\partial x}$$ (2.53)

where $\epsilon_e$ is the volume fraction of liquid and the current density $j$ accounts for the incoming or outgoing flow of lithium due to the redox reactions. Plugging in the expression for the ion flux, $N$, gives the desired equation for the time rate of change of lithium concentration in the liquid.

43
Concentrated solution theory is more generalized, and therefore more accurate, since it accounts for all interactions between species. For concentrated solutions, the Nernst-Einstein equation no longer holds \[67\], and a different approach must be taken to obtain the ion flux. Rather than separate out the effects from diffusion and migration, consider that both phenomena are ultimately driven by a gradient in the electrochemical potential of species in the fluid. The electrochemical potential of a species, $\bar{\mu}_i$ (J/mol), combines its chemical potential, $\mu_i$ (J/mol), with the influence of an external electrostatic potential, $\Phi$ (V):

$$
\bar{\mu}_i = \mu_i + z_i F \Phi
$$

(2.54)

where $z_i$ is the charge number of species $i$. Note that an uncharged species ($z_i = 0$) will not be affected by the electrostatic potential, in which case $\bar{\mu}_i = \mu_i$.

The chemical potential of a species is a measure of how much the Gibbs free energy of the system will change with respect to a change in the number of moles of that species. In other words, it is equal to the slope of the Gibbs free energy curve, $\mu_i = \frac{\partial G}{\partial n_i}$, at a constant temperature, pressure, and number of moles of other species \[68\]. The process of diffusion occurs in the presence of a concentration gradient because species are trying to minimize the Gibbs free energy of the system, and therefore move from areas of high chemical potential to areas of low chemical potential.

Consider the schematic in Fig. 2.6, where the concentration of species A in region 1 is greater than in region 2 (assume no electrostatic potential). In region 1, the Gibbs free energy is higher than is possible for the system, and the chemical potential of species A, $\mu_A^1$, has a positive value. In region 2, the Gibbs free energy is also higher than is possible, but the chemical potential of species A, $\mu_A^2$, has a negative value. Molecules moving out of region 1 will shift point 1 to the left on the curve, thereby decreasing $G$. Conversely, molecules moving into region 2 will shift point 2
Figure 2.6: Schematic showing the differences in chemical potential between region 1 and 2, and the opportunity to lower the Gibbs free energy via diffusion.

to the right on the curve, also decreasing $G$. For this reason, molecules diffuse from region 1 to 2 (high to low concentration), and stop when the chemical potentials of species in each region are equal and $G$ is minimized. Generally speaking, diffusion will occur along a concentration gradient from areas of high chemical potential to areas of low chemical potential; however, this is not always the case, like for example in the presence of an opposing electric field.

Migration is also a process that is driven by differences in electrochemical potential, only it occurs for charged particles in the presence of an electric field, which generates an electrostatic potential. Consider the schematic in Fig. 2.7, where a positively charged and negatively charged particle are subject to an electric field.
Figure 2.7: Schematic showing the differences in electrostatic potential between position 1 and 2. Particles will move toward regions of lower electrochemical potential.

The electrostatic potential ($\Phi = -\nabla E$) at position 1 is greater than at position 2. Therefore, according to Eq. (2.54), the positively charged particle ($z_i > 0$) will reach a region of lower electrochemical potential by moving towards position 2. The negatively charged particle ($z_i < 0$) will reach a region of lower electrochemical potential by moving towards position 1.

For a concentrated solution, the flux in Eq. (2.52) can be replaced with the multicomponent diffusion equation, involving the electrochemical potential of each species [68]:

$$c_i \frac{\partial \bar{\mu}_i}{\partial x} = \sum_j K_{ij} (v_j - v_i)$$

(2.55)

where $\mu_i$ (J/mol) is the electrochemical potential of species $i$, $K_{ij}$ (Js/m$^5$) is the friction (or interaction) coefficient between species $i$ and $j$, and $v_i$ (m/s) is the velocity of species $i$. This equation essentially combines the effects of diffusion and migration by describing how a gradient in electrochemical potential induces species to move.
The friction coefficient can be redefined as a function of the species concentration:

$$K_{ij} = \frac{\bar{R}Tc_i c_j}{c_T \bar{D}_{ij}} \quad (2.56)$$

where the total $c_T = \sum_i c_i$ is the sum of all species concentrations, and $\bar{D}$ is a diffusion coefficient describing the interaction of species $i$ and $j$ (considered to be an empirical parameter). For Li-ion batteries, a typical electrolyte contains a lithium salt that breaks up into two ions in the solvent, such as $LiPF_6 \rightarrow Li^+ + PF_6^-$. In this case, species $i = +, -, 0$, indicating either the cation (+), anion (-), or solvent (0). For this two ion system, Eq. (2.55) for each species becomes:

$$c_+ \frac{\partial \mu_+}{\partial x} = K_{+-}(v_- - v_+) + K_{+0}(v_0 - v_+) \quad (2.57)$$

$$c_- \frac{\partial \mu_-}{\partial x} = K_{-+}(v_+ - v_-) + K_{-0}(v_0 - v_-) \quad (2.58)$$

$$c_0 \frac{\partial \mu_0}{\partial x} = K_{0+}(v_+ - v_0) + K_{0-}(v_- - v_0) \quad (2.59)$$

Newton’s third law of motion requires that $K_{ij} = K_{ji}$ and $\bar{D}_{ij} = \bar{D}_{ji}$ (reactions are equal and opposite), meaning that the sum of the right hand sides of Eqs. (2.57)-(2.59) equals zero. The left hand sides also sum to zero by the Gibbs-Duhem relation [68]. Therefore, only 2 of these 3 equations are independent and the third equation for the solvent can be eliminated (the focus here is on the ions). Applying the definition of a liquid phase current density $i_e = F \sum_i z_i N_i = F \sum_i z_i c_i v_i$, Eqs. (2.57)-(2.58) can be rearranged to solve for the ionic fluxes:

$$N_+ = c_+ v_+ = -\frac{\nu_e \bar{D}_e c_T}{\nu_e RT c_0} \frac{\partial \mu_e}{\partial x} + \frac{i_e(t) t^+_0}{z_+ F} + c_+ v_0 \quad (2.60)$$

$$N_- = c_- v_- = -\frac{\nu_e \bar{D}_e c_T}{\nu_e RT c_0} \frac{\partial \mu_e}{\partial x} + \frac{i_e(t) t^-_0}{z_- F} + c_- v_0 \quad (2.61)$$
where a number of substitutions have been made. First, $\nu_+$ and $\nu_-$ are the number of cations and anions produced by the dissociation of one molecule of salt (both equal to 1 in the case of LiPF$_6$). Properties of the total salt are defined as: $\nu_e = \nu_+ + \nu_-$, $\bar{\mu}_e = \nu_+ \bar{\mu}_+ + \nu_- \bar{\mu}_-$, and $c_e = \frac{c_+}{\nu_+} = \frac{c_-}{\nu_-}$. The definition of the salt electrolyte concentration, $c_e$, comes from the assumption of electroneutrality. An overall thermodynamic diffusion coefficient for the electrolyte, based on the gradient in chemical potential as the driving force, can be defined as:

$$\bar{D}_e = \frac{\bar{D}_0^+ \bar{D}_0^- (z_+ - z_-)}{z_+ D_{0+} - z_- D_{0-}}$$ (2.62)

This electrolyte diffusion coefficient essentially represents a compromise between the diffusion coefficients of the anion and cation (the ions effectively diffuse together at the same rate, due to electroneutrality). Finally, the transference numbers, $t^+_i$, represent the fraction of current carried by species $i$, defined as:

$$t^+_0 = 1 - t^-_0 = \frac{z_+ \bar{D}_{0+}}{z_+ D_{0+} - z_- D_{0-}}$$ (2.63)

The Li-ion flux in Eq. (2.60) could now be applied to Eq. (2.53) to obtain the time rate of change of Li-ions; however, two more modifications are useful. First, the salt diffusion coefficient that is usually measured and reported, $D_e$, is based on a concentration gradient as the driving force, rather than a gradient in chemical potential, but is related to $\bar{D}_e$ by:

$$D_e = \bar{D}_e \frac{c_0}{c_T} \left( 1 + \frac{d \ln \gamma_{+-}}{d \ln m} \right)$$ (2.64)

where $\gamma_{+-}$ is the mean molal activity coefficient and $m$ is the number of moles of electrolyte per kilogram of solvent. Second, the dependence on the gradient of chemical
potential can be replaced by a dependence on a concentration gradient according to:

\[ \frac{D_e}{\nu_eRT} \frac{c_T}{c_0} \frac{\partial \bar{\mu}_e}{\partial x} = D_e \left( 1 - \frac{d \ln c_0}{d \ln c_e} \right) \frac{\partial c_e}{\partial x} \tag{2.65} \]

The lithium ion flux now becomes:

\[ N_+ = -D_e \left( 1 - \frac{d \ln c_0}{d \ln c_e} \right) \frac{\partial c_e}{\partial x} + \frac{i_e(t)}{t_0^+} \frac{c_e}{z_+ \nu_+ F} + c_0 \nu_0 \tag{2.66} \]

Assuming no bulk motion of the solvent (no convection), \( \nu_0 = 0 \), and Eq. (2.66) can be substituted into Eq. (2.53) to obtain the time rate of change of lithium ions:

\[ \epsilon_e \frac{\partial c_e}{\partial t} = a_i j_i(x,t) F - \frac{\partial}{\partial x} \left[ -D_e \left( 1 - \frac{d \ln c_0}{d \ln c_e} \right) \frac{\partial c_e}{\partial x} + \frac{i_e(t)}{t_0^+} \frac{c_e}{z_+ \nu_+ F} \right] \tag{2.67} \]

By assuming the change in solvent concentration for a given change in salt concentration is close to zero, i.e. \( \frac{d \ln c_0}{d \ln c_e} \approx 0 \), Eq. (2.67) can be simplified to:

\[ \epsilon_e \frac{\partial c_e}{\partial t} = a_i j_i(x,t) F + D_e \frac{\partial^2 c_e}{\partial x^2} - \frac{\partial i_e}{\partial x} \frac{t_0^+}{z_+ \nu_+ F} \tag{2.68} \]

Finally, applying Eq. (2.2), which gives the relationship between liquid phase current density and the transfer current density in electrode \( i \), i.e. \( a_i j_i(x,t) = -\frac{\partial i_e}{\partial x} \), the final expression is:

\[ \epsilon_e \frac{\partial c_e}{\partial t} = D_e \frac{\partial^2 c_e}{\partial x^2} + \frac{3 \epsilon_i(1-t_0^+)}{F R_i} j_i(x,t) \tag{2.69} \]

where the transfer current density term is positive in the negative electrode, negative in the positive electrode, and zero in the separator.

Equation (2.69) is applied separately in each electrode and the separator, where the three equations are coupled at the region boundaries, so that the concen-
tration and concentration flux are matched:

\[
\begin{align*}
\frac{\partial c_e}{\partial x} \bigg|_{x=L_n^-} &= \frac{\partial c_e}{\partial x} \bigg|_{x=L_n^+}, & c_e \bigg|_{x=(L_n+L_{sep})^-} &= c_e \bigg|_{x=(L_n+L_{sep})^+} \\
\frac{\partial c_e}{\partial x} \bigg|_{x=L_n^-} &= \frac{\partial c_e}{\partial x} \bigg|_{x=L_n^+}, & \frac{\partial c_e}{\partial x} \bigg|_{x=(L_n+L_{sep})^-} &= \frac{\partial c_e}{\partial x} \bigg|_{x=(L_n+L_{sep})^+}
\end{align*}
\]

(2.70)

where the superscripts − and + are used to indicate approaching from the left and right, respectively. Here the diffusion coefficient \( D_e \) is treated as constant in all three regions.

### 2.4.8 Potential in the liquid

In order to develop an expression for a measurable potential drop across the electrolyte, two reference electrodes are required. If the reactions at the reference electrodes are sufficiently fast, the measured potential drop across the two reference electrodes is equal to the potential drop across the electrolyte. Consider two pure lithium reference electrodes immersed in electrolyte at two different positions, where there may be a gradient in the ion concentrations (and therefore electrochemical potentials) between the two positions. The goal is to find an expression for the potential drop between the two reference electrodes. The reaction at the reference electrodes involves the cation, anion, and possibly the solvent:

\[
s_--M_i^{z-} + s_+M_i^{z+} + s_0M_0 \rightleftharpoons n_e e^-
\]

(2.71)

where \( M_i \) represents the chemical formula of species \( i \), \( s_i \) is the stoichiometric coefficient, and \( n_e \) is the number of electrons on the products side. The electrostatic potential gradient between the two reference electrodes is related to the electrochemi-
ical potential gradients of each species:

\[
\begin{align*}
\frac{s_-}{\partial x} \frac{\partial \bar{\mu}_-}{\partial x} + \frac{s_+}{\partial x} \frac{\partial \bar{\mu}_+}{\partial x} + s_0 \frac{\partial \bar{\mu}_0}{\partial x} &= -n_e F \frac{\partial \phi_e}{\partial x} \tag{2.72}
\end{align*}
\]

It is desirable to express Eq. (2.72) in terms of the overall salt electrochemical potential gradient, \(\frac{\partial \bar{\mu}_e}{\partial x}\). To do this, the first two terms in Eq. (2.72) can be expressed as (recalling that \(s_+ z_+ + s_- z_- = -n\) and \(\bar{\mu}_e = \nu_+ \bar{\mu}_+ + \nu_- \bar{\mu}_-\)):

\[
\frac{s_-}{\partial x} \frac{\partial \bar{\mu}_-}{\partial x} + \frac{s_+}{\partial x} \frac{\partial \bar{\mu}_+}{\partial x} = \frac{s_+}{\nu_+} \frac{\partial \bar{\mu}_e}{\partial x} - \frac{n_e}{z_-} \frac{\partial \bar{\mu}_-}{\partial x} \tag{2.73}
\]

Using the Gibbs-Duhem equation, \(c_0 d\bar{\mu}_0 + c_e d\bar{\mu}_e = 0\), the third term in Eq. (2.72) can be replaced with:

\[
\frac{s_0}{\partial x} \frac{\partial \bar{\mu}_0}{\partial x} = -\frac{s_0}{\partial x} c_e \frac{\partial \bar{\mu}_e}{\partial x} \tag{2.74}
\]

After making these substitutions, equation (2.72) becomes:

\[
\left( \frac{s_+}{n_e \nu_+} - \frac{s_0 c_e}{n_e c_0} \right) \frac{\partial \bar{\mu}_e}{\partial x} - \frac{1}{z_-} \frac{\partial \bar{\mu}_-}{\partial x} = -F \frac{\partial \phi_e}{\partial x} \tag{2.75}
\]

Finally, to remove the \(\frac{\partial \bar{\mu}_-}{\partial x}\) term, the following relation is used:

\[
\frac{1}{z_-} \frac{\partial \bar{\mu}_-}{\partial x} = -\frac{F}{\sigma_e} i_e(t) - \frac{t_0^+}{z_+ \nu_+} \frac{\partial \bar{\mu}_e}{\partial x} \tag{2.76}
\]

which is obtained by substituting the ion fluxes in Eq. (2.60)-(2.61) into Eq. (2.55). Here the conductivity, \(\sigma_e\), is defined by:

\[
\frac{1}{\sigma_e} = \frac{-\bar{R}T}{c_T z_+ z_- F^2} \left( \frac{1}{D_{+\bar{\mu}}} + \frac{c_0 t_0}{c_+ D_{0-}} \right) \tag{2.77}
\]
The result is that the electrostatic potential gradient can be expressed in terms of the salt electrochemical potential gradient and the liquid current density:

$$\frac{\partial \phi_e}{\partial x} = \frac{i_e(t)}{\sigma_e} - \frac{1}{F} \left( \frac{s_+}{n_e v_+} + \frac{t_0^+}{z_+ v_+} - \frac{s_0 c_e}{n c_0} \right) \frac{\partial \bar{\mu}_e}{\partial x}$$

(2.78)

Once again, it is useful for the model to express the electrochemical potential gradient as a concentration gradient, using the relationships in Eqs. (2.64)-(2.65):

$$\frac{\partial \phi_e}{\partial x} = \frac{i_e(t)}{\sigma_e} - \frac{\nu_e \bar{R} T}{F} \left( \frac{s_+}{n_e v_+} + \frac{t_0^+}{z_+ v_+} - \frac{s_0 c_e}{n c_0} \right) \left( 1 + \frac{d \ln \gamma_{+e}}{d \ln \gamma_e} \right) \frac{1}{c_e} \frac{\partial c_e}{\partial x}$$

(2.79)

Again, the assumption has been made that $\frac{d \ln \rho_0}{d \ln c_e} \approx 0$. For the pure lithium reference electrode, the reaction only involves the cation, i.e. $Li^+ + e^- \rightleftharpoons Li$, so that $\nu_+ = 1$, $\nu_e = 2$, $z_+ = 1$, $s_+ = 1$, $s_0 = s_0 = 0$, and $n_e = -1$. This simplifies the expression to:

$$\frac{\partial \phi_e}{\partial x} = \frac{i_e(t)}{\sigma_e} + \frac{2 \bar{R} T (1 - t_0^+)}{F} (1 + \gamma) \frac{\partial \ln (c_e)}{\partial x}$$

(2.80)

where $\gamma = \frac{d \ln \gamma_{+e}}{d \ln \gamma_e}$ represents the effective activity coefficient. The potential gradient in the liquid therefore consists of an ohmic potential drop, due to a resistance to ion flow, and a potential drop due to the concentration gradient.

2.4.9 Solution method

In total, the 1+1D porous electrode model consists of 2 partial differential equations (PDEs) in each of the solid and liquid phases, one governing lithium concentrations and one governing potential distributions. The PDEs of each phase are coupled by the transfer current density, $j(x,t)$, which is defined by the charge transfer overpotential at the electrode-electrolyte interface. The most common solution approach is to discretize the two spatial coordinates and time, and solve for the concentrations and
potentials at each node using a finite difference or finite volume method. For example, a finite difference approach approximates first and second derivatives of concentration at a given node $j$ as:

$$\frac{\partial c}{\partial x} \bigg|_j \approx \frac{c_{j+1} - c_{j-1}}{2\Delta x}$$

(2.81)

$$\frac{\partial^2 c}{\partial x^2} \bigg|_j \approx \frac{c_{j+1} - 2c_j - c_{j-1}}{\Delta x^2}$$

(2.82)

Applying these approximations at each discretized node, results in $N$ equations to solve for $N$ unknown concentrations. Care must be taken to properly set the time step size and spatial step size so that the solution remains stable. In short, for a given spatial step size, the time step size must be sufficiently small.

### 2.5 The Single Particle Model

The $1 + 1$D porous electrode model can be further simplified by eliminating the dependence of concentrations and potentials on the coordinate $x$, within each electrode. Instead, a single spherical particle can be considered to represent the solid phase in each electrode. The transfer current density is then averaged over the entire electrode active surface area, for each electrode $i = n, p$:

$$j_i(x, t) = j_i(t) = \frac{R_i I(t)}{3AL_i \epsilon_i}$$

(2.83)

where $R$ (m) is the particle radius, $A$ (m$^2$) is the current collector area, $L$ (m) is the electrode thickness, and $\epsilon$ is the volume fraction of active material (the substitution $a_i = \frac{3\epsilon_i}{R_i}$ has been made for the spherical shape). This approach is called a uniform utilization model, since it considers the transfer current density to be uniform over the entire electrode. A transfer current density that is uniform for each electrode allows direct integration of the solid conduction Eq. (2.50) across the thickness of a
given electrode, resulting in a potential drop that is proportional to the current (i.e. a purely Ohmic relationship). The proportionality constant is the ohmic resistance for that electrode, defined as $R_{c,i}$.

Removing the dependence on $x$ reduces the computational complexity considerably, and is often considered a minimum requirement for developing an electrochemical model simple to be used for online estimation. Another common approximation is to also neglect the concentration gradient in the electrolyte and simply treat the potential drop across the electrolyte to be Ohmic, i.e. proportional to the current. This is a good approximation at low current rates ($< 1C$). A model that uses the uniform utilization assumption along with the removal of non-Ohmic liquid dynamics is often referred to in the literature as a single particle model (SPM). The model used in this work does not neglect these non-Ohmic liquid dynamics (instead it approximates them using model order reduction), but it is still referred to as a single particle model (technically it could be called an “enhanced” SPM).

### 2.6 Model Order Reduction

Often in dynamic system modeling applications, particularly when online control or estimation is required, simulating the original governing model equations is too computationally expensive. In this case, it is helpful to use model order reduction (MOR) techniques to reduce the complexity of the original model. This essentially involves reducing the number of states required to represent the dynamic system of interest. Of course the trade-off is that the reduced order model is less accurate; however, if done intelligently, loss of accuracy over the expected operating conditions can be minimized, while still obtaining significant improvements in computational efficiency.

Despite the simplifications using the single particle approximation, the model still
relies on a PDE governing lithium diffusion in each electrode particle (Eq. (2.48)) and a PDE governing lithium diffusion and migration in the electrolyte (Eq. (2.69)). However, a system of ordinary differential equations (ODEs), rather than PDEs, is desired for traditional control and estimation algorithms. One approach could be to discretize the spatial coordinates using a finite difference method. The resulting set of finite difference equations can be treated as a set of discrete state equations, where the concentration at each node becomes a state. The major limitation of this approach is that it typically requires a large number of states to accurately represent the concentration profile. Alternatively the number of states, and therefore the computational complexity, can be reduced by employing various model order reduction techniques. In this section, MOR techniques are analyzed for the solid and liquid diffusion PDEs.

The following additional variables are used in the following sections on MOR and state space formulation

- $A_i$: State space matrix $A$ for sub-model $i$ ()
- $B$: Integration constant ()
- $B_i$: State space matrix $B$ for sub-model $i$ ()
- $C_s(s)$: Laplace transform of the surface concentration (mol/m$^3$)
- $\bar{C}(s)$: Laplace transform of the bulk concentration (mol/m$^3$)
- $G_0(s)$: Original nonlinear transfer function ()
- $G_x(s)$: Approximated linear transfer function ()
- $a_{j,i}$: Numerator coefficients of the approximated transfer functions ()
- $b_{j,i}$: Denominator coefficients of the approximated transfer functions ()
- $m$: Polynomial approximation coefficients ()
- $s$: Complex frequency used in Laplace transform ()
- $u$: Input to the state space model ()
- $x_i$: State vector of sub-model $i$ ()
2.6.1 Solid diffusion equation

The diffusion of lithium in a given solid particle is governed by Fick’s law in spherical coordinates:

\[
\frac{\partial c_i}{\partial t} = D_i \frac{r^2}{\partial r} \left( r^2 \frac{\partial c_i}{\partial r} \right),
\]

\[
\frac{\partial c_i}{\partial r} \bigg|_{r=0} = 0, \quad D_i \frac{\partial c_i}{\partial r} \bigg|_{r=R_i} = \pm \frac{j_i(t)}{F}
\]

where the current density (Eq. (2.83)) defines the flux at the particle radius, using the uniform utilization approximation. To obtain the concentration profile as a function of time and the radial direction \( r \), a numerical approach is taken, using a finite difference approximation. The PDE is discretized in time from 0 to \( t_{end} \) (step size of \( \Delta t = 0.1s \)), and space from 0 to \( R_i \) (step size of \( \Delta r = 0.02\mu m \)). Figure 2.8 shows representative concentration profiles at different points in time for a positive electrode particle under a 1C discharge current. However, the output equation of the single particle model only depends on the concentration at the surface of the particle. This provides an opportunity for model order reduction, to obtain a relationship just between the surface concentration and the input current, while sacrificing knowledge about the exact shape of the profile.

**Polynomial approximation**

As is evident from Fig. 2.8, the shape of the concentration profile appears to be somewhat parabolic. It is therefore possible to model the profile as some polynomial function of \( r \). Typically, a second order polynomial is used:

\[
c_i(r) = m_{0,i} + m_{1,i}r + m_{2,i}r^2
\]
Figure 2.8: Representative concentration profiles along the radial distance $r$ at various times during a 1C discharge.

The idea of the polynomial approximation is to relate the surface concentration to the bulk concentration, which has simple first order dynamics, using the assumed parabolic shape. The bulk concentration dynamics are calculated by accounting for the current in and out of the spherical particle, normalized appropriately to convert from current to concentration:

$$\frac{\partial \bar{c}_i}{\partial t} = \frac{I(t)}{FAL_i\epsilon_i}$$  \hspace{1cm} (2.86)

To solve for the polynomial coefficients $m_{0,i}$, $m_{1,i}$, and $m_{2,i}$, consider the boundary conditions: $\frac{\partial c_i}{\partial r}$ at $r = 0$ and $\frac{\partial c_i}{\partial r} = \frac{j_i(t)}{D_i F} = \frac{i(t)R_i}{3D_i FAL_i\epsilon_i}$ at $r = R_i$, where $\frac{\partial c_i}{\partial r} = m_{1,i} + 2m_{2,i}r$. Finally, the volume averaged concentration for the assumed parabolic shape must be equal to the bulk concentration $\bar{c}$:

$$\bar{c}_i = \int_0^{R_i} c_i(r) \frac{4\pi r^2}{\frac{4}{3}\pi R_i^3} dr$$  \hspace{1cm} (2.87)
where the term $\frac{4\pi r^2}{3\pi R_i}$ is needed to account for the spherical coordinates. Using the two boundary conditions and Eq. (2.87), the polynomial coefficients are found to be:

$$m_{0,i} = \bar{c}_i - \frac{R_i^2 I(t)}{10D_i FAL_i \epsilon_i}, \quad m_{1,i} = 0, \quad m_{2,i} = \frac{I(t)}{6D_i FAL_i \epsilon_i}$$ (2.88)  

Evaluating the concentration profile at the particle surface, $r = R_i$, gives a relation between surface and bulk concentration:

$$c_{s,i} = \bar{c}_i + \frac{R_i^2 I(t)}{15D_i FAL_i \epsilon_i}$$ (2.89)

Note that the deviation from the bulk concentration is proportional to the current: a higher current results in a larger deviation. The procedure for simulating the model is to first calculate the bulk concentration using Eq. (2.86), calculate the surface concentration using Eq. (2.89), and then evaluate the OCV and charge transfer overpotential using the surface concentration.

To obtain the transfer function relating surface or bulk concentration to current density, the Laplace transform of Eqs. (2.86) and (2.89) become:

$$s \bar{C}_i(s) = \frac{1}{FAL_i \epsilon_i} I(s)$$ (2.90)  

$$C_{s,i}(s) = \bar{C}_i(s) + \frac{R_i^2}{15D_i FAL_i \epsilon_i} I(s)$$ (2.91)

Rearranging and combining terms gives the transfer functions:

$$\frac{\bar{C}_i(s)}{I(s)} = \frac{1}{FAL_i \epsilon_i s}$$ (2.92)  

$$\frac{C_{s,i}(s)}{I(s)} = \frac{15D_i + R_i^2 s}{15D_i FAL_i \epsilon_i s}$$ (2.93)
**Padé approximation**

The Padé approximation is a model order reduction method applied in the frequency domain, in order to approximate a high order or possibly nonlinear transfer function with a lower order linear transfer function. The idea is to apply a technique called “moment matching” to match the approximated linear transfer function and its derivatives to the original transfer function and its derivatives. The matching must be done at some specified frequency; therefore, typically the approximated linear transfer function will match well near the specified frequency, and start to deviate as the frequency increases or decreases. The order of the desired approximated linear transfer function also affects the accuracy of the approximation. Higher order approximations allow higher derivatives to be matched at the specified frequency, which allows the approximation to remain accurate over a larger frequency range. To apply the Padé approximation to the solid diffusion equation, a transcendental transfer function is obtained by taking the Laplace transform of Eq. (2.84) [35, 60]:

\[
sC_i(s) = D_i \frac{d^2 C_i(s)}{dr^2} + \frac{2D_i}{r} \frac{dC_i(s)}{dr},
\]

\[
\frac{dC_i(s)}{dr} = 0 \text{ at } r = 0, \quad D_i \frac{dC_i(s)}{dr} = \mp \frac{R_i I(s)}{3FA L_i \epsilon_i} \text{ at } r = R_i
\] (2.94)

By treating \(s\) as a parameter, Eq. 2.94 becomes an ODE with the solution (using the symbolic manipulation software Maple):

\[
C_i(s,r) = \frac{B_1}{r} \exp \left( r \sqrt{\frac{s}{D_i}} \right) + \frac{B_2}{r} \exp \left( -r \sqrt{\frac{s}{D_i}} \right)
\] (2.95)

Applying the boundary conditions at \(r = 0\) and \(r = R_i\) to solve for integration constants \(B_1\) and \(B_2\), and evaluating the result at the particle surface \(r = R_i\) gives
the transfer function between surface concentration and current:

\[
\frac{C_{s,i}(s)}{I(s)} = G_0(s) = \frac{R_i^2}{3D_iFAL_i\varepsilon_i} \frac{\sinh \left( \sqrt{\frac{sR_i}{D_i}} \right)}{\sqrt{\frac{sR_i}{D_i}} R_i \cosh \left( \sqrt{\frac{sR_i}{D_i}} \right) - \sinh \left( \sqrt{\frac{sR_i}{D_i}} \right)}
\]  

(2.96)

This is the transfer function that will be approximated as linear, using moment matching. The form of the approximated linear transfer (i.e. the ratio of zeros to poles) can be selected depending on the application. For the solid diffusion equation, a good approximated transfer function, \( G_x(s) \), is in the form:

\[
\frac{C_{s,i}(s)}{I(s)} \approx G_x(s) = a_{0,i} + a_{1,i}s + a_{2,i}s^2 + \ldots \]

\[
\frac{C_{s,i}(s)}{I(s)} \approx G_x(s) = a_{0,i} + a_{1,i}s + a_{2,i}s^2 + \ldots \]

(2.97)

where the pole at \( s = 0 \) is added, recognizing the fact that the intercalation of lithium in the solid is inherently an integration process. Moment matching involves differentiating both the transcendental and approximated linear transfer functions with respect to \( s \) and evaluating the limits as \( s \) goes to the specified frequency of choice. Then the two transfer functions and their respective derivatives are equated, in order to solve for the coefficients \( a_{j,i} \) and \( b_{j,i} \). For the diffusion equation, the primary goal is to capture the steady state and low frequency behavior, so the limits are taken as \( s \to 0 \). Therefore, the moment matching procedure results in a system of \( m \) equations, where \( m \) is the number of unknown coefficients in (2.97):

\[
\begin{align*}
\lim_{s \to 0} G_0(s) &= \lim_{s \to 0} G_x(s) \\
\lim_{s \to 0} \frac{dG_0(s)}{ds} &= \lim_{s \to 0} \frac{dG_x(s)}{ds} \\
\lim_{s \to 0} \frac{d^2G_0(s)}{ds^2} &= \lim_{s \to 0} \frac{d^2G_x(s)}{ds^2} \\
\vdots \\
\lim_{s \to 0} \frac{d^{m-1}G_0(s)}{ds^{m-1}} &= \lim_{s \to 0} \frac{d^{m-1}G_x(s)}{ds^{m-1}}
\end{align*}
\]

(2.98)
The resulting linear transfer function coefficients up to a 3rd order approximation are shown in Table 2.1. Note that the first order approximation simply amounts to integrating the appropriately scaled current (the same result as in Eq. (2.86)), and so in this case the bulk and approximated surface concentrations are the same.

**Method comparison**

Figure 2.9 shows the step response of the various approximations compared to the numerical solution of the PDE in the time domain. For this simulation, a positive electrode particle is subject to a 1C discharge pulse for 30s, followed by a 30s rest. Overall the 2nd and 3rd order Padé approximations are much more accurate than the polynomial approximation, particularly during the time immediately after the current is applied or removed. While all the approximations converge to the correct steady state concentration at the end of the pulse or at the end of the rest, the polynomial approximation is not able to capture any dynamics associated with the rest and simply reverts back to the bulk concentration value as soon as the current is stopped. However, the tradeoff with the polynomial approximation is that it only requires one
Figure 2.9: Step response of the surface concentration for various reduced order models, compared to the original PDE numerical solution.

model state, while an $n$th order Padé approximation requires $n$ states. A comparison is also done in the frequency domain in Fig. 2.10, by using the transcendental transfer function in Eq. (2.96) as the true transfer function of the PDE. In general the approximations are accurate at low frequency (this should certainly be true for the Padé approximation since it was specifically designed to match at $s = 0$), and diverge at higher frequency. In [60], an analysis was done to determine an appropriate Padé approximation order based on the expected frequency composition of a typical drive cycle. A 3rd order approximation was deemed sufficient to accurately capture the dynamics for most of the frequencies seen in the US06 highway and Federal Urban Dynamometer Schedule (FUDS) cycles. Therefore, a 3rd order Padé approximation of the solid diffusion equation is used in the model in this work. It is interesting to note that all approximations remain accurate up to a higher frequency for a smaller particle size or a larger diffusion coefficient. A smaller particle or faster diffusion dynamics
Figure 2.10: Frequency response of the surface concentration for various reduced order models, compared to the original PDE transcendental transfer function.

effectively cause a steady state to be reached sooner, allowing the approximations to remain accurate up to higher frequencies.

2.6.2 Liquid diffusion equation

The diffusion of lithium ions in the electrolyte is given by Fick's law in Cartesian coordinates, with a source term added to account for the influx of lithium into or out of the electrolyte:

$$\epsilon_e \frac{\partial c_e}{\partial t} = D_e \frac{\partial^2 c_e}{\partial x^2} \pm \frac{3(1-t_i^+)}{F R_i} j_i(t) \quad (2.99)$$

Using the uniform utilization assumption, Eq. (2.99) is solved separately in each region (negative electrode, separator, and positive electrode), where the source term has a positive value in the negative electrode (for discharge), is zero in the separator,
and has a negative value in the positive electrode (for discharge). The three solutions are coupled at the region boundaries, so that the concentration and concentration flux are matched:

\[ c_e|_{x=L_n^-} = c_e|_{x=L_n^+}, \quad \frac{\partial c_e}{\partial x}|_{x=L_n^-} = \frac{\partial c_e}{\partial x}|_{x=L_n^+}, \quad \frac{\partial c_e}{\partial x}|_{x=(L_n+L_{sep})^-} = \frac{\partial c_e}{\partial x}|_{x=(L_n+L_{sep})^+} \] (2.100)

where the superscripts − and + are used to indicate approaching from the left and right, respectively. Again, the liquid diffusion coefficient is assumed to be constant in all regions. Additional zero flux boundary conditions are required at the electrode ends:

\[ \left. \frac{\partial c_e}{\partial x} \right|_{x=0} = 0, \quad \left. \frac{\partial c_e}{\partial x} \right|_{x=L_{cell}} = 0 \] (2.101)

The result is three 2nd order PDEs with six total boundary conditions. Again, to obtain the concentration profile as a function of time, the set of PDEs is solved numerically using a finite difference approach (\( \Delta t = 0.05s \) and \( \Delta x = 3.625\mu m \)). Representative profiles under a 1C discharge are shown in Fig. 2.11 at different points in time. Note that the positive electrode is thicker than the negative, since positive electrode materials generally have a lower energy density which must be compensated for with a thicker electrode. To model the cell voltage dynamics, only the potential drop across the liquid is required (the difference between the potential at \( x = 0 \) and \( x = L_{cell} \)). This also means that only the liquid concentrations at the electrode ends are required to compute the potential difference via Eq. (2.80). Once again, this provides an opportunity to use model order reduction to relate the concentration at the electrode ends to the input current, while sacrificing knowledge of the entire profile as a function of \( x \).
Figure 2.11: Representative liquid concentration profiles along the x direction, at various times for a 1C discharge.

**Padé approximation**

The same process for obtaining the transcendental transfer function can be used for the liquid diffusion equation. After taking the Laplace transform of Eq. (2.99), the solution has the form (again using Maple):

\[
C_{e,i}(s) = B_{1,i} \exp \left( \frac{\epsilon_e s}{D_e} x + B_{2,i} \exp - \sqrt{\frac{\epsilon_e s}{D_e}} x \pm \frac{(1 - t_0^+)}{FAL_i \epsilon_i \epsilon_e} I(s) \right)
\]

where the input term is added in the negative electrode, subtracted in the positive electrode, and zero in the separator. The integration constants \(B_{1,i}\) and \(B_{2,i}\) in each region are then solved by applying the boundary conditions. The resulting transcendental transfer functions relating concentration at each electrode end to input current are obtained by evaluating at \(x = 0\) and \(x = L_{cell}\). To simplify the expressions, the
following assumptions about the component thicknesses are made, that are typical of a lithium-ion cell: \( L_n = \frac{1}{4} L_{cell} \), \( L_{sep} = \frac{3}{20} L_{cell} \), \( L_p = \frac{3}{5} L_{cell} \). Using these assumptions, the approximated transfer functions for the concentration at each electrode end are:

\[
\frac{C_{e,n}(s)}{I(s)} = -\frac{5(1 - t_0^+) \sinh \left( \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right)}{6 \epsilon_e F A L_{cell} \epsilon_n s \cosh \left( \frac{2}{5} \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right) \sinh \left( \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right)} - \frac{5(1 - t_0^+) \sinh \left( \frac{1}{5} \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right)}{6 \epsilon_e F A L_{cell} \epsilon_n s \cosh \left( \frac{2}{5} \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right) \sinh \left( \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right)} + \frac{12(1 - t_0^+) \sinh \left( \frac{3}{5} \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right)}{6 \epsilon_e F A L_{cell} \epsilon_n s \cosh \left( \frac{2}{5} \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right) \sinh \left( \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right)} + \frac{12(1 - t_0^+) \sinh \left( \frac{7}{5} \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right)}{6 \epsilon_e F A L_{cell} \epsilon_n s \cosh \left( \frac{2}{5} \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right) \sinh \left( \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right)} + \frac{5(1 - t_0^+) \sinh \left( \frac{2}{5} \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right)}{6 \epsilon_e F A L_{cell} \epsilon_n s \cosh \left( \frac{2}{5} \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right) \sinh \left( \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right)}
\]

\( C_{e,p}(s) \) is calculated similarly:

\[
\frac{C_{e,p}(s)}{I(s)} = -\frac{5(1 - t_0^+) \sinh \left( \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right)}{3 \epsilon_e F A L_{cell} \epsilon_n s \sinh \left( \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right)} + \frac{12(1 - t_0^+) \sinh \left( \frac{1}{4} \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right)}{3 \epsilon_e F A L_{cell} \epsilon_n s \sinh \left( \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right)} + \frac{5(1 - t_0^+) \sinh \left( \frac{2}{3} \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right)}{3 \epsilon_e F A L_{cell} \epsilon_n s \sinh \left( \sqrt{\frac{\epsilon e}{D_e}} L_{cell} \right)}
\]
Table 2.2: Linear transfer function coefficients for liquid diffusion using a Padé approximation of various orders.

<table>
<thead>
<tr>
<th>Order</th>
<th>$C_{e,n}(s)$</th>
<th>$I(s)$</th>
<th>$a_{0,n}$</th>
<th>$a_{1,n} s + b_{2,n} s^2 + \ldots$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>$\frac{(1-t_0^n)}{FA e_n} \frac{1}{1+b_{2,n} s}$</td>
<td>$a_{0,n}$</td>
<td>$0.33 \frac{L_{cell}}{D_e}$</td>
<td>$0.33 \frac{L_{cell}}{D_e}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$a_{1,n}$</td>
<td></td>
<td>$0.012 \frac{L_{cell}^3}{D_e^2}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$b_{2,n}$</td>
<td>$0.092 \frac{L_{cell}^2}{D_e}$</td>
<td>$0.13 \frac{L_{cell}^2}{D_e}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$b_{3,n}$</td>
<td></td>
<td>$0.0027 \frac{L_{cell}}{D_e}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Order</th>
<th>$C_{e,p}(s)$</th>
<th>$I(s)$</th>
<th>$a_{0,p}$</th>
<th>$a_{1,p} s + a_{2,p} s^2 + \ldots$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>$\frac{(1-t_0^p)}{FA e_p} \frac{1}{1+b_{2,p} s}$</td>
<td>$a_{0,p}$</td>
<td>$-0.25 \frac{L_{cell}}{D_e}$</td>
<td>$-0.25 \frac{L_{cell}}{D_e}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$a_{1,p}$</td>
<td></td>
<td>$-0.0045 \frac{L_{cell}^3}{D_e^2}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$b_{2,p}$</td>
<td>$0.11 \frac{L_{cell}^2}{D_e}$</td>
<td>$0.13 \frac{L_{cell}^2}{D_e}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$b_{3,p}$</td>
<td></td>
<td>$0.0029 \frac{L_{cell}}{D_e}$</td>
</tr>
</tbody>
</table>

Unlike the solid diffusion equation, the approximated linear transfer function in the liquid does not require the pole at $s = 0$:

$$\frac{C_{e,i}(s)}{I(s)} \approx \frac{a_{0,i} + a_{1,i} s + a_{2,i} s^2 + \ldots}{1 + b_{2,i} s + b_{3,i} s^2 + \ldots} \quad (2.105)$$

Performing the moment matching procedure gives the coefficients, as shown in Table 2.2.

Figure 2.12 shows the step response of the Padé approximations compared to the numerical solution of the PDE in the time domain. Again, the simulations considers a 1C discharge pulse for 30s, followed by a 30s rest. It is evident that even the 1st order approximation matches the PDE very well. A comparison of the frequency response is shown in Fig. 2.13 for the negative electrode (the positive electrode shows a similar result). Once again, even the 1st order approximation matches the PDE
Figure 2.12: Step response of the liquid concentration for various Pade approximation orders, compared to the original PDE numerical solution.

over most of the frequency range. Clearly, the liquid diffusion dynamics are much closer to a first order response than the solid diffusion dynamics. Therefore, a 1st order Padé approximation for liquid diffusion is used in the model in this work. Note that although it is possible to implement a polynomial approximation similar to the approach taken for solid diffusion, it is not necessary in this case. The 1st order Padé approximation is more accurate while adding only adding 1 state per electrode.

2.7 Reduced Order Model in State Space

For the online state estimation done in this work, the selected model must have a relatively low number of states but also maintain good accuracy at high current rates and moderately high frequencies (~1 Hz). Therefore, the base model used in this work uses the uniform utilization assumption, the Padé approximation to relate
the particle surface concentration to the input current using a set of ODEs, but also includes approximated liquid dynamics to improve accuracy at high rates. The model is further modified to accommodate a composite electrode, i.e. a single electrode with more than one active material mixed together. A composite electrode is treated as two particles, acting in parallel, one representing each active material. The parallel arrangement assures that the potential across each parallel path is identical, but the current is split between each path according to its effective impedance. The fraction of current going to each parallel path is defined as \( \beta_i \). Rather than outline the model for a generic set of single or composite electrodes, from here on the negative electrode is considered to be a single material graphite (C) electrode and the positive electrode is considered to be a composite electrode of LiMn\(_2\)O\(_4\) (LMO) and LiNi\(_{1/3}\)Mn\(_{1/3}\)Co\(_{1/3}\)O\(_2\) (NMC). The graphite-composite LMO/NMC battery is a common chemistry used in...
plug-in and battery electric vehicles on the road today, and is the chemistry of the cells used to experimentally validate the SOC/SOH estimation algorithms in this work. However, the model could easily be modified to include any combination of single or composite material electrodes by adding or removing parallel paths. Figure 2.14 shows the model structure for the graphite-LMO/NMC couple. Subscripts $n$ and $p$ are used to refer to the negative and positive electrodes, respectively, while subscripts $LMO$ and $NMC$ are used referring specifically to one active material in the composite electrode. Subscript $i$ indicates a generic particle, where $i = n, p, LMO, or NMC$.

The state space representation of this model for both electrode and electrolyte dynamics is outlined in this section. In the model, current is considered to be the input and cell voltage the output (galvanostatic control), since this is the most common approach in electrified vehicle applications. Alternatively, one could treat the voltage as the input and current as the output (potentiostatic control).

### 2.7.1 Electrode state space dynamics

For each representative electrode particle, there are sources of concentration, charge transfer, and ohmic overpotential. The resulting electrode potential deviates from
the OCV after taking into account each source of overpotential:

\[
\Phi_i(t) = U_i(c_{s,i}(t)) - \eta_i(c_{s,i}(t), \beta_i(t)I(t)) - R_{c,i}\beta_i(t)I(t) \tag{2.106}
\]

The parameter \(\beta_i(t)\) is the fraction of current going to the given active material. For the single material negative electrode, \(\beta_n(t) = 1\); however, for the composite positive electrode, \(\beta_{LMO}(t)\) and \(\beta_{NMC}(t)\) vary dynamically and are calculated at each time step, as shown in Section 4.2.2.

The charge transfer overpotential associated with the redox reaction at the surface of the particle is obtained by inverting the Butler-Volmer law:

\[
\eta_i(t) = \frac{2RT}{F} \sinh^{-1} \left( \frac{j_i(t)}{2i_0,i(t)} \right) \tag{2.107}
\]

where \(i_0\) is the exchange current is a function of the lithium surface concentration in the solid and the local lithium concentration in the electrolyte:

\[
Fk_i\sqrt{c_{s,i}(t)c_{e,i}(t)(c_{max,i} - c_{s,i}(t))} \tag{2.108}
\]

The uniform utilization assumption is used to obtain the current density for each active material:

\[
j_i(t) = \frac{\beta_i(t)I(t)}{AL_i\alpha_i\epsilon_i} \tag{2.109}
\]

Note that for each material within the composite electrode, the active material volume fraction \(\epsilon_i\) incorporates not just the overall ratio of active to inactive material, but also the material split of LMO to NMC. In other words, \(\epsilon_{LMO}\) represents the ratio of LMO active material volume to all other material volume, including NMC and inactive material. This concept will be reintroduced in Chapter 5, when attempting to estimate active material loss from the composite electrode.
The concentration overpotential is the result of the concentration gradient that forms throughout an electrode particle, as lithium diffuses from the interior of the particle to the reaction zone at the surface (or vice versa). The concentration throughout a spherical particle is given by Fick's law of mass diffusion, in spherical coordinates (Eq. (2.48)).

Finally, an ohmic overpotential arises as electrons flow from the particle surfaces to the current collector and through the weld joints to the external battery tabs. Again, since the current density is assumed to be spatially constant within each electrode, this ohmic overpotential is simply equal to \( R_{c,i} \beta_i(t) I(t) \).

In the solid phase, the important metrics for tracking cell SOC and available power are the particle bulk concentration and particle surface concentration, respectively. By taking the Laplace transform of Eq. (2.86), a linear transfer function (pure integrator) is obtained relating the bulk concentration to the current input.

\[
\frac{\bar{C}_i(s)}{I(s)} = \frac{\beta_i \epsilon_i s}{FAL_i \epsilon_i s} \quad (2.110)
\]

where the current split factor, \( \beta_i \), dictates the fraction of current applied to a particle in the composite electrode (for the negative electrode, \( \beta_n = 1 \)). In state space, the bulk concentration dynamics are:

\[
\dot{\bar{x}}_i = \bar{c}_i = \frac{\beta_i}{FAL_i \epsilon_i} u \quad (2.111)
\]

where \( u \) is the input current.

Alternatively, a 3rd order Pade approximation of Eq. (2.48), evaluated at the particle radius, is used to obtain a linear transfer function between the surface con-
centration and the current input $[35, 60]$.

$$\frac{C_{s,i}(s)}{I(s)} = \frac{\beta_i R_i}{3FAL_i \epsilon_i} \frac{a_{0,i} + a_{1,i}s + a_{2,i}s^2}{s(1 + b_{2,i}s + b_{3,i}s^2)}$$  \hspace{1cm} (2.112)

where $a_{0,i} = \frac{3}{R_i}$, $a_{1,i} = \frac{4R_i}{11D_i}$, $a_{2,i} = \frac{R_i^3}{105D_i}$, $b_{2,i} = \frac{3R_i^2}{55D_i}$, $b_{3,i} = \frac{R_i^4}{3465D_i}$. The electrode particle sub-model can now be formulated in state space, where the surface concentrations can be written as linear combinations of the 3 states, and the input is current:

$$\dot{x}_i = A_i x_i + B_i(t)u$$  \hspace{1cm} (2.113)

$$A_i = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -\frac{1}{b_{3,i}} & -\frac{b_{2,i}}{b_{3,i}} \end{bmatrix}, \quad B_i(t) = \begin{bmatrix} 0 \\ 0 \\ \mp \left( \frac{\beta_i(t)R_i}{3FAL_i \epsilon_i} \right) \frac{1}{b_{3,i}} \end{bmatrix}$$

$$c_{s,i} = \begin{bmatrix} a_{0,i} & a_{1,i} & a_{2,i} \end{bmatrix} x_i$$

The sign applied in $B_i$ is negative for the negative electrode and positive for the positive electrode.

Both the bulk and surface concentrations are of interest, but tracking both separately would require 4 total states per particle. However, it is possible to write the bulk concentration as a function of the 3 states from the Padé approximation, through some algebraic manipulation $[35]$. This is possible due to the fact that both bulk and surface concentration transfer functions contain a pure integrator, $\frac{1}{s}$. In this way, estimating 3 states gives both the surface and bulk concentrations. First, use the fact that the same input current appears in both Eqs. (2.111) and the third state equation in (2.113), and equate the two expressions (a positive electrode particle is used here):

$$\dot{x}_i = \dot{c}_i = \frac{\beta_i}{FAL_i \epsilon_i} u = \frac{3}{R_i} b_{3,i} \left( x_{3,i} + \frac{1}{b_{3,i}} x_{2,i} + \frac{b_{2,i}}{b_{3,i}} x_{3,i} \right)$$  \hspace{1cm} (2.114)
This implies that the dynamics of the bulk concentration are related to the dynamics of the 3 Padé states by:

\[
\dot{\bar{c}}_i = \begin{bmatrix} a_{0,i} & a_{0,i}b_2 & a_{0,i}b_3,i \end{bmatrix} \dot{x}_i \tag{2.115}
\]

where the substitutions \( a_{0,i} = \frac{3}{R_i} \), \( x_{2,i} = \dot{x}_{1,i} \), and \( x_{3,i} = \dot{x}_{2,i} \) have been made. Since the bulk and surface concentrations are equal at steady state, it follows that:

\[
\bar{c}_i = \begin{bmatrix} a_{0,i} & a_{0,i}b_2 & a_{0,i}b_3,i \end{bmatrix} x_i \tag{2.116}
\]

Therefore, the complete state space representation, including the output potential is:

\[
\dot{x}_i = A_i x_i + B_i(t)u \tag{2.117}
\]

\[
A_i = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -\frac{1}{b_{3,i}} & -\frac{b_{2,i}}{b_{3,i}} \end{bmatrix}, \quad B_i(t) = \begin{bmatrix} 0 \\ 0 \\ \frac{\beta_i(t)R_i}{3FAL_{ii}} \end{bmatrix}
\]

\[
\begin{bmatrix} c_{s,i} \\ \bar{c}_i \end{bmatrix} = \begin{bmatrix} a_{0,i} & a_{1,i} & a_{2,i} \\ a_{0,i} & a_{0,i}b_2,i & a_{0,i}b_3,i \end{bmatrix} x_i
\]

\[
\phi_i = U_i(c_{s,i}) \pm \eta_i(c_{s,i}, \beta_i(t)u) \pm R_{c,i}\beta_i(t)u \tag{2.118}
\]

where the output electrode potential is a nonlinear function of the states and input.

The result of this substitution is that both bulk and surface concentration can be written as linear combinations of the same 3 states. In other words, there are 6 total states in the composite positive electrode sub-model and 3 total states in the single material negative electrode sub-model. The states are updated in discrete time using a finite difference solution.
2.7.2 Electrolyte state space dynamics

In the liquid phase, the concentration and potential are governed by:

\[
\epsilon_e \frac{\partial c_e}{\partial t} = D_e \frac{\partial^2 c_e}{\partial x^2} \pm \frac{3(1 - t_0^+)}{F R_i} j_i(t) \quad (2.119)
\]

\[
\frac{\partial \phi_e}{\partial x} = \frac{i_e(t)}{\sigma_e} + \frac{2RT(1 - t_0^+)}{F} (1 + \gamma) \frac{\partial \ln(c_e)}{\partial x} \quad (2.120)
\]

The first thing to note is that in terms of its effect on the overall cell voltage, only the potential difference across the electrolyte (from one electrode to the other) is important. So while \(c_e\) varies continuously in the \(x\) direction, the only locations where the concentration effects the liquid potential difference is at the two electrode ends. With this in mind, it is only required to evaluate the concentration at the electrode ends. A similar situation occurs for the solid diffusion problem in the spherical particles, where only the surface concentration is of interest, rather than the exact concentration profile in the radial direction. Once again, this provides an opportunity to use the Padé approximation to relate the concentration at the point of interest to the input current. In this case, a first order Padé approximation of Eq. (2.119) is sufficient to relate the lithium concentration at each electrode end to the input current (or current density). The resulting first order transfer functions relate liquid phase concentration at each electrode end \((x = 0\) and \(x = L_{cell}\)) to the current input:

\[
\frac{C_{e,n}(s)}{I(s)} = \frac{(1 - t_0^+)}{F A \epsilon_n} \frac{a_{0,n}}{1 + b_{2,n}s} \quad (2.121)
\]

\[
\frac{C_{e,p}(s)}{I(s)} = \frac{(1 - t_0^+)}{F A \epsilon_p} \frac{a_{0,p}}{1 + b_{2,p}s} \quad (2.122)
\]

where \(a_{0,n} = 0.33 \frac{L_{cell}}{D_e}, \quad b_{2,n} = 0.092 \epsilon_e \frac{L_{cell}^2}{D_e}, \quad a_{0,p} = -0.25 \frac{L_{cell}}{D_e}, \quad b_{2,p} = 0.11 \frac{\epsilon_e L_{cell}^2}{D_e} \).

Once concentrations at each electrode are obtained, the liquid potential equation
can be integrated directly from one electrode end to the other. However, the ionic current density, $i_e$, still needs to be defined as a function of $x$. The uniform utilization assumption dictates that the transfer current density between solid and liquid phase is constant for a given electrode. Recall Eq. (2.2) for a 1 dimensional problem and a transfer current density independent of $x$:

$$a_i j_i(t) = \frac{\partial i_s}{\partial x} = -\frac{\partial i_e}{\partial x}$$  \hspace{1cm} (2.123)$$

The spatially constant $j$ means that within a given electrode, at a given time, the slope of $i_e$ (and $i_s$) with respect to $x$ is constant. Applying the logical boundary conditions that $i_e = 0$ at the current collectors and $i_s = 0$ in the separator gives the picture of the solid and liquid phase current density shown in Fig. 2.15.

After defining the liquid phase current density profile over $x$, Eq. (2.120) can be integrated from $x = 0$ to $x = L_{cell}$, with the lithium concentration at each electrode assumed to be a known constant (obtained from the Pade approximation). This gives the liquid voltage drop:

$$V_e(t) = \phi_e(x = 0) - \phi_e(x = L_{cell})$$

$$= \frac{I(t)(\frac{1}{2}L_p + L_{sep} + \frac{1}{2}L_n)}{\sigma_e A} + \frac{2\bar{R}T(1 - \bar{t}_0^+)}{F} \ln \left( \frac{\bar{c}_{e,n}(t)}{\bar{c}_{e,p}(t)} \right)$$ 

$$\hspace{1cm} (2.124)$$
In state space the liquid concentrations are proportional to the states, the input is current, and the output is a nonlinear function of the states and input.

\[
\dot{x}_e = A_e x_e + B_e u \tag{2.125}
\]

\[
A_e = \begin{bmatrix}
\frac{1}{b_{1,e,p}} & 0 \\
0 & \frac{1}{b_{1,e,n}}
\end{bmatrix} \quad B_e = \begin{bmatrix}
\frac{1}{b_{2,e,p}} \\
\frac{1}{b_{2,e,n}}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\bar{c}_{e,p} \\
\bar{c}_{e,n}
\end{bmatrix} = \begin{bmatrix}
1 & 1
\end{bmatrix} x_e
\]

\[
V_e = u \left( \frac{\frac{1}{2}L_p + L_{sep} + \frac{1}{2}L_n}{\sigma_e A} \right) + \frac{2\bar{R}T(1 - t_0^+)}{F} (1 + \gamma) \ln \left( \frac{\bar{c}_{e,n}}{\bar{c}_{e,p}} \right) \tag{2.126}
\]

### 2.7.3 Cell voltage output

Finally the cell voltage output (which is the quantity actually being measured), combines the outputs from the electrode and electrolyte sub-models, plus an extra cell level Ohmic resistance loss associated with the flow of electrons through the current collectors to the battery tabs:

\[
V_{cell} = \Phi_p(c_{e,p}, u) - \Phi_n(c_{e,n}, u) - V_e(c_{e,p}, c_{e,n}, u) - R_{cell} u \tag{2.127}
\]

In total, the complete battery model contains 3 states representing the bulk and surface concentrations of each material in the composite positive electrode (3 for LMO and 3 for NMC), 3 states representing the bulk and surface concentrations in the negative electrode, and 2 states representing the Li-ion concentrations in the electrolyte at the positive and negative electrode ends.
2.8 Electrochemical Models for State Estimation

The goal of this dissertation is to use the state space model outlined in this chapter to estimate model parameters associated with SOC and SOH. Traditionally, SOC/SOH estimation has been done by applying state estimation techniques with the use of equivalent circuit models. However, electrochemical models used for SOC/SOH estimation offer several advantages over traditional circuit models. First, electrochemical-based models can more accurately predict the cell input-output behavior by considering nonlinear overpotentials. In [77], a single particle model (neglecting liquid dynamics) was compared to a R-RC circuit model and found to give a better prediction of the cell voltage under constant current. The authors used the polynomial approximation to obtain the electrode surface concentration, which resulted in similar computation times between the SPM and circuit model. It is likely that with a more dynamic current profile the SPM would perform even better in comparison, since the nonlinear effects of diffusion and charge transfer would become more important.

Second, electrochemical models are generally easier to parameterize and require fewer experiments to do so. Circuit models must rely on resistances and capacitances that are scheduled based on SOC, current, and/or temperature. This requires a large number of experiments varying SOC, current, and temperature to construct lookup tables between circuit parameters and each metric. On the other hand, the parameters in electrochemical models have physical meaning. Many of these parameters, like electrode area, thickness, porosity, and active material volume, are directly measurable (offline) for a given cell. Most electrochemical model parameters do not require scheduling based on SOC or current like in a circuit model, as these nonlinear effects are directly accounted for by considering diffusion and charge transfer overpotentials. Some parameters do vary with temperature, like rate constants and conductivity,
but they generally follow the well-known Arrhenius relationship, and so require fewer experiments to parameterize.

Third, electrochemical models that incorporate physically meaningful parameters have the potential to facilitate more aggressive and safe use of the battery. Rather than enforce cell level voltage limits to ensure safe operation, limits could be placed on the relevant electrochemical parameters that drive the unwanted phenomena that occur outside the voltage limits. For example, rather than enforce an upper voltage limit to avoid lithium plating at the anode, a limit could be placed on the negative electrode overpotential, ensuring that it remains > 0 [90]. Alternatively, rather than enforce a lower voltage limit to avoid overdepleting and damaging the anode, a limit could be placed on the surface concentration of the electrodes [90]. In certain situations, enforcing these new limits could allow the traditional voltage limits to be violated without damaging the cell. In other cases, the voltage limits may not be conservative enough, and enforcing limits on physical parameters could prevent damage from occurring.

Finally, and most importantly in the context of SOH estimation and prognostics, electrochemical models contain parameters that are directly associated with degradation mechanisms. For example, loss of electrode active material either through dissolution or binder cracking is associated with a decrease in the active material volume fraction, $\epsilon_i$, as active material dissolves or becomes electrically isolated from the current collector. The electrode conductivity, $\sigma_i$, may decrease as well, as contact between particles becomes diminished. Alternatively, loss of cyclable Li due to SEI growth is associated with a decrease in the total lithium content at any given time (equal to the sum of the lithium in each electrode, $\bar{c}_n c_{\text{max},n} A L_n \epsilon_n + \bar{c}_p c_{\text{max},p} A L_p \epsilon_p$). SEI growth may also cause an increase in the charge transfer overpotential, as the SEI layer acts as a barrier to lithium reaching the reaction site [4].

Using an electrochemical model allows these physical parameters, and their ef-
fects on cell performance, to be estimated directly. This is a major improvement over a circuit model, which can only estimate the overall capacity loss or resistance rise [23,34,36,55,71–73,80,99]. For example, Fig. 2.16 shows a model prediction of the capacity loss for a composite LMO-NMC cell under four different aging scenarios. In each scenario 20% of the capacity is lost; however, the different degradation mechanisms result in very different discharge curves. In this case, only estimating the total capacity loss using a circuit model is not enough to capture the complex changes in the voltage curve, and the accuracy of the voltage prediction would decrease as the cell ages. On the other hand, by estimating the parameters associated with each mechanism using an electrochemical model, a more accurate model can be maintained over the life of the battery. An accurate model allows for an accurate estimate of SOC, which is essential in order for a vehicle’s BMS to function properly. Distinguishing between degradation mechanisms also allows for a better estimate of available power. For example the power fade associated with the loss of LMO active material would be greater than for loss of NMC (LMO is the “high power” material due to its faster kinetics). Using electrochemical models also improves the possibilities for prognostics, as model parameters can be linked to physical degradation models which can predict how the parameters might evolve and affect battery performance in the future.

The advantages of using electrochemical models for state estimation are clear; however, there are also several disadvantages. First, electrochemical models are generally going to require a higher number of states than a circuit model. For the most simplified SPM (1st order electrode dynamics and neglected liquid dynamics), the computational effort is similar to a circuit model [77]; however to get the most benefit out of the electrochemical model (particularly at high rates), higher order electrode dynamics and at least 1st order liquid dynamics may be required. In this case, the computational effort would be higher than for a typical circuit model. Computational
Figure 2.16: Model prediction of a 1C discharge curve under different degradation mechanisms, each resulting in 20% capacity fade.

effort is clearly a concern for onboard state estimation.

Second, although some of the electrochemical model parameters are straightforward to measure, others like diffusion coefficients and kinetic rate constants are much less certain. For instance, the relaxation times and subsequent diffusion coefficients assessed via the potentiostatic or galvanostatic intermittent titration technique (PITT or GITT), or electrochemical impedance spectroscopy (EIS), can vary by orders of magnitude over the range of lithium composition [37, 61]. This has been reported for prominent active materials such as layered oxides like NMC [87, 97]. Kinetic rate constants are typically extracted from Tafel or EIS experiments and exhibit similar variability amongst measurement methods [89]. Fortunately, SOC estimator performance is fairly robust to uncertainty in diffusion and kinetic parameters [61].

Finally, an electrochemical model that contains separate states and parameters for each electrode may have problems with the observability of the model, since only
the cell voltage can be realistically measured in an automotive application. This issue of observability will be discussed in Chapter 3, but essentially, it can be difficult to uniquely identify all of the states and parameters of interest using only the cell voltage measurement. These problems with observability are avoided in circuit models that do not try to separate out the dynamics of each electrode.

2.9 Summary

A wide range of battery models exist, differing in the length scales considered and in the computational complexity. However, online model-based estimation algorithms must rely on models that are computationally cheap, while still providing meaningful information. Porous electrode theory describes the concentration and potential variation in the solid and liquid phases, but the reliance on partial differential equations makes these models ill-suited for control and estimation applications. The single particle approximation reduces the computational complexity by neglecting the spatial variation of current density through-the-thickness of the cell, and considering a single representative particle for each electrode active material. The concentration gradient within the representative particles or liquid phase can be approximated further by using model order reduction methods like the polynomial or Padé approximation to convert the diffusion PDEs into low order ODEs that can be represented by transfer functions or in state space. To accommodate the composite electrode, the traditional single particle model structure is extended to include two parallel pathways for each composite active material, where the current is split between each path but the potential across each path is the same.

Finally, electrochemical models have distinct advantages over traditional circuit models, including the ability to retain physically meaningful parameters that are associated with SOC and SOH. That said, electrochemical models attempt to provide
additional information about the battery without any new measurements, and so problems arise with how well the states and parameters of interest can be uniquely identified.
Chapter 3

Methods for State Estimation

In Chapter 2, a reduced-order electrochemical model is developed for a composite electrode lithium-ion cell. The goal in this dissertation is to use the electrochemical model to design state estimation algorithms, in order to estimate model states or parameters associated with SOC and SOH. This chapter provides an overview of concepts related to estimation, that are ultimately utilized to design the SOC and SOH estimation algorithms described in Chapters 4 and 5.

3.1 Introduction to Estimation

Broadly speaking, a dynamic system is comprised of inputs, outputs, and internal states. Commanded inputs and outputs are measured by sensors (current and voltage, for example), which are subject to noise but are considered to be known quantities at each time step. The internal states are variables that completely characterize the system at any given time step. States are defined so that no knowledge of the past is needed to predict their evolution from one time step to the next. In other words, knowledge of the states at one time step, along with the input, is enough to predict the states at the next step. State variables can be physical quantities like concentration,
temperature, or position, or they can have no physical meaning but mathematically represent the dynamics of, say, a reduced-order model. States that have physical meaning may or may not be directly measurable. Lithium concentrations, for example, are not practically measurable in a battery pack onboard a vehicle, but they are useful quantities for a BMS to know. Determining the model states from input and output measurements can be done under three scenarios, referred to as prediction, estimation, and smoothing [96]. In prediction, the states at some time step $k$ are determined using only past measurements, i.e. a prediction of the future state. In estimation, the states at time $k$ are determined using past and present measurements, up to and including time $k$. In smoothing, the states at time $k$ are determined using past, present, and future measurements. Needless to say, in smoothing there is a delay in when the states can be determined, since it requires waiting until future measurements are available. Although there are certainly reasons why prediction of SOC and/or SOH would be of interest, the focus of this work is on estimation and smoothing. Essentially the goal is to determine the SOC and SOH at the present time, or possibly in the near past. For determining SOC, only estimation is applicable since SOC estimates must be made in real time. Real time knowledge of SOC has implications in determining how much instantaneous battery power is available and in determining the optimal ICE/EM power split at each time step. On the other hand, smoothing becomes an option for determining SOH, since the parameters associated with SOH only change significantly over the course of months or years, and a delay in obtaining the next estimate may be acceptable.

This chapter outlines the fundamental concepts behind observability and observer design, with a focus on nonlinear systems. Observability of the single particle model used in this work is discussed. Finally, specific observer types including the extended Kalman filter, Kalman smoother, and particle filter are presented.

The following variable definitions are assumed in this chapter.
A  Linear state space matrix A
B  Linear state space matrix B
C  Linear state space matrix C
D  Linear state space matrix D
I_d  Identity matrix
K  Kalman gain
L_f  Lie derivative
P  Estimate error covariance matrix
Q  Process noise covariance matrix
R  Sensor noise covariance matrix
e  Estimate error
f  Nonlinear state dynamics function
g  Nonlinear output function
n  Number of states
p  Probability distribution
s_i  Cubic spline coefficient for the electrode OCV
t  Time
u  Input to the state space model
x  Model state vector
y  Model output
κ  Observability matrix condition number
\mathcal{O}  Linear observability matrix
\mathcal{O}'  Nonlinear observability matrix
Subscript d  Indicates discrete time
Subscript k  Discrete time step
Superscript −  Indicates a priori
Superscript +  Indicates a posteriori

3.2 Observability

When estimating unmeasurable states based on measured outputs, it is easy to imagine a situation in which the number of measured outputs is not sufficient to uniquely identify the states. For example, take the ideal gas relationship between pressure, \( P \), volume, \( V \), number of moles, \( n \), and temperature, \( T \): \( PV = nRT \). Consider pressure, volume, and temperature to be the states of the system. However if, say, pressure is the only available measurement, then the system cannot be uniquely defined. There are an infinite number of \( V \) and \( T \) combinations that result in the same measured
pressure. In this case, it can be said that the states of this system are not observable from the output, which introduces the concept of observability. The states of a system are observable if they can be uniquely reconstructed using measurements of the output over a finite period of time \([26, 39, 45, 88]\). Since the set of state variables at any given time step completely characterize the system, the concept of observability is often defined as the ability to reconstruct the initial state vector, \(x_0\). If this initial condition is known, then in theory the entire state vector trajectory over time is also known (assuming knowledge of the model and input).

### 3.2.1 Observability for linear systems

For a linear time invariant (LTI) system, the state dynamics, \(\dot{x}(t)\), input, \(u(t)\), and output, \(y(t)\), are assumed to have the form:

\[
\dot{x}(t) = Ax(t) + Bu(t) \tag{3.1}
\]

\[
y(t) = Cx(t) + Du(t) \tag{3.2}
\]

When dealing with real systems with data acquisition systems that take samples at discrete time intervals, it is helpful to convert the continuous time system dynamics to discrete time using a finite difference approximation at time step \(k\):

\[
\dot{x}_k \approx \frac{x_k - x_{k-1}}{\Delta t} = Ax_{k-1} + Bu_{k-1} \tag{3.3}
\]

Solving for \(x_k\) gives:

\[
x_k = (I_d + \Delta tA)x_{k-1} + \Delta tBu_{k-1} = A_dx_{k-1} + B_du_{k-1} \tag{3.4}
\]
where $I$ is the identity matrix. This particular finite difference approximation is first-order accurate, meaning that the truncation error due to the discretization is proportional to the time step size.

The output is sampled at each time step:

$$y_k = Cx_k + Du_k$$

Consider a single-input, single-output system. The states of the system are observable if the initial state vector, $x_0$, can be uniquely reconstructed from samples of the output. If there is more than one state, more than one sample of the output must be recorded to have any chance at reconstructing $x_0$. In fact, at least $n$ samples must be recorded in order to identify $n$ states in the initial $x_0$ (results in $n$ equations for $n$ unknowns). Consider the first 3 samples of the output for a 3 state system, at time steps $k = 0, 1, 2$:

$$y_0 = Cx_0 + Du_0$$
$$y_1 = Cx_1 + Du_1$$
$$y_2 = Cx_2 + Du_2$$

Since the goal is to calculate $x_0$, the state dynamics in Eq. (3.4) can be used to relate the state vectors $x_1$ and $x_2$ to the state vectors at the previous time step:

$$y_0 = Cx_0 + Du_0$$
$$y_1 = C(A_dx_0 + B_du_0) + Du_1$$
$$y_2 = C(A_dx_1 + B_du_1) + Du_2$$
Repeating this again for the third equation gives:

\[ y_0 = Cx_0 + Du_0 \]
\[ y_1 = C(A_d x_0 + B_d u_0) + Du_1 \] \hspace{1cm} (3.8)
\[ y_2 = C(A_d(A_d x_0 + B_d u_0) + B_d u_1) + Du_2 \]

Isolating \( x_0 \) in the system of equations gives the matrix equation:

\[
\begin{bmatrix}
y_0 - Du_0 \\
y_1 - Du_1 - CB_d u_0 \\
y_2 - Du_2 - CB_d u_1 - CA_d B_d u_0
\end{bmatrix} =
\begin{bmatrix}
C \\
CA_d \\
CA_d^2
\end{bmatrix} x_0
\]

\hspace{1cm} (3.9)

\[ z = M x_0 \]

Determining a unique solution to this matrix equation, i.e. \( x_0 = M^{-1} z \), requires that the matrix \( M \) is invertible. There are several tests to determine if a matrix is invertible, one of which is to calculate the matrix rank. A square matrix is invertible if its rank is equal to the dimension of the matrix (3 in this case). This gives rise to the well-known rank test for a linear system with \( n \) states, which defines an observability matrix:

\[
O = \begin{bmatrix}
C \\
CA_d \\
\vdots \\
CA_d^{n-1}
\end{bmatrix}
\]

\hspace{1cm} (3.10)

The states of the system are observable if: \( rank(O) = n \). Note that system observability does not depend on the input (\( B_d \) or \( D \) matrices).

A similar process can be done to test observability for the original LTI system in continuous time, Eqs. (3.1-3.2). However, instead of gaining more information by looking at \( n \) consecutive samples of the output, more information can be gained
for the continuous time system by considering the output along with its first \( n - 1 \) derivatives with respect to time. Again, consider the case where \( n = 3 \):

\[
\begin{align*}
y(t) &= Cx(t) + Du(t) \\
\dot{y}(t) &= C\dot{x}(t) + D\dot{u}(t) \\
\ddot{y}(t) &= C\ddot{x}(t) + D\ddot{u}(t) = C\frac{d}{dt}\dot{x}(t) + D\ddot{u}(t) \\
\end{align*}
\]

(3.11)

Substituting in the state dynamics gives:

\[
\begin{align*}
y(t) &= Cx(t) + Du(t) \\
\dot{y}(t) &= C(Ax(t) + Bu(t)) + D\dot{u}(t) = CAx(t) + CBu(t) + D\dot{u}(t) \\
\ddot{y}(t) &= C\frac{d}{dt}(Ax(t) + Bu(t)) + D\ddot{u}(t) = CA\dot{x}(t) + CB\dot{u}(t) + D\ddot{u}(t) \\
&= CA^2x(t) + CABu(t) + CB\dot{u}(t) + D\ddot{u}(t) \\
\end{align*}
\]

(3.12)

Once again, the linear system allows separation of the components that are dependent on \( x(t) \), leading to the matrix equation:

\[
\begin{bmatrix}
y(t) - Du(t) \\
\dot{y}(t) - D\dot{u}(t) - CBu(t) \\
\ddot{y}(t) - D\ddot{u}(t) - CB\dot{u}(t) - CABu(t)
\end{bmatrix} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} x(t)
\]

(3.13)

with the same observability matrix as in the discrete time case:

\[
O = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}
\]

(3.14)

which must be full rank in order to solve for \( x(t) \) via matrix inversion.
3.2.2 Observability for nonlinear systems

For a general nonlinear dynamic system, the state dynamics and outputs cannot be written as a matrix equation, and are instead written as generic functions of \( x \) and \( u \):

\[
\dot{x}(t) = f(x(t), u(t)) \tag{3.15}
\]
\[
y(t) = g(x(t), u(t)) \tag{3.16}
\]

Here, the assumption is that \( f \) and \( g \) are \( C^\infty \) (or smooth) functions, with derivatives of any order that exist and are continuous. When dealing with a nonlinear system, the question of whether the state vector has a unique solution from the output and its derivatives becomes less clear. However, a similar approach to the linear case in continuous time can be taken by analyzing the time derivatives of the output. Consider a single-input, single-output system (from here on the dependence on \( t \) is dropped in the notation). Using the chain rule, the first time derivative of the output is:

\[
\dot{y} = g_x(x, u) \dot{x} + g_u(x, u) \dot{u} \\
= g_x(x, u) f(x, u) + g_u(x, u) \dot{u} \tag{3.17}
\]

where the notation \( g_i \) is used to indicate the Jacobian of the function \( g \) with respect to the variable \( i \). For the single-input, single-output case, these just become: \( g_x = \begin{bmatrix} \frac{\partial g}{\partial x_1} & \cdots & \frac{\partial g}{\partial x_n} \end{bmatrix} \) and \( g_u = \frac{\partial g}{\partial u} \). Differentiating Eq. (3.17) with respect to time again
gives the second time derivative of the output:

\[ \ddot{y} = \dot{g}_x(x, u, \dot{u}) \dot{x} + \dot{g}_u(x, u, \dot{u}) \dot{u} + \ddot{g}_u(x, u, \dot{u}) \ddot{u} \]

\[ = \ddot{g}(x, u, \dot{u}, \ddot{u}) \]

(3.18)

At this point, the term “Lie derivative” of the function \( g \) is often used, simply defined as a given time derivative of the output. Repeated Lie derivatives are denoted as:

\[ L_f^0(g) = y = g(x, u) \]

\[ L_f^1(g) = \dot{y} = g_x(x, u)f(x, u) + g_u(x, u) \dot{u} \]

\[ L_f^2(g) = \ddot{y} = g_x(x, u, \dot{u})f(x, u) + g_u(x, u, \dot{u}) \dot{u} + \ddot{g}_u(x, u, \dot{u}) \ddot{u} \]

\[ : \]

(3.19)

In a general sense, the output and its \( n - 1 \) derivatives can be written in the form:

\[
\begin{bmatrix}
  y \\
  \dot{y} \\
  \vdots \\
  y^{(n-1)}
\end{bmatrix} =
\begin{bmatrix}
  g(x, u) \\
  \dot{g}(x, u, \dot{u}) \\
  \vdots \\
  g^{(n-1)}(x, u, \dot{u}, \ldots, u^{(n-1)})
\end{bmatrix} =
\begin{bmatrix}
  L_f^0(g) \\
  L_f^1(g) \\
  \vdots \\
  L_f^{n-1}(g)
\end{bmatrix}
\]

(3.20)

Together, the output and its \( n - 1 \) derivatives form a nonlinear system of \( n \) equations, where the objective is to solve for the \( n \) components in the state vector \( x \). The ability to find a unique solution to this system of equations implies the states are observable.

For a general nonlinear system, there is no simple test that can determine if a unique solution exists for any output \( y \). However, the requirement can be weakened by only testing for a solution \( x \) that is unique within some neighborhood of states around \( x \).

To do this, consider a general set of equations that satisfy \( h(x, y) = 0 \), where the
number of equations is equal to the number of components in $x$. The implicit function theorem \[40\] states that for some nominal values $x_0$ and $y_0$, such that $h(x_0, y_0) = 0$, if the Jacobian matrix with respect to $x$:

$$\frac{\partial h}{\partial x} = \begin{bmatrix}
\frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial h_n}{\partial x_1} & \cdots & \frac{\partial h_n}{\partial x_n}
\end{bmatrix}$$  \hspace{1cm} (3.21)

is nonsingular at $x_0$ and $y_0$, then for all $y$ close to $y_0$, the function has a solution $x$ that is locally unique. Applying this theorem to the system of equations in Eq. (3.20), a locally unique solution of the state vector $x$ exists if the Jacobian of the set of Lie derivatives:

$$\mathcal{O}' = \begin{bmatrix}
\frac{\partial L_0^0(g)}{\partial x_1} & \cdots & \frac{\partial L_0^0(g)}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial L_{n-1}^n(g)}{\partial x_1} & \cdots & \frac{\partial L_{n-1}^n(g)}{\partial x_n}
\end{bmatrix}$$  \hspace{1cm} (3.22)

is nonsingular, i.e. $\text{rank}(\mathcal{O}') = n$. Formally, a system that satisfies this rank condition is referred to as locally weakly observable. The qualifiers locally and weakly have specific mathematical definitions \[13\]. The system 3.15-3.16 is weakly observable at $x_0$ if there exists a neighborhood $U$ of $x_0$ such that there is no indistinguishable state from $x_0$ in $U$. Two state pairs $x_0$ and $x_0'$ are indistinguishable if for any $u$ and any $t \geq 0$, $g(x(t, u, x_0), u) = g(x(t, u, x_0'), u)$, where $x(t, u, x_0)$ is the state solution from $t_0$ to $t$, initialized at $x_0$, and subject to input $u$. Essentially, two state pairs are indistinguishable if the outputs that they produce are identical. A weakly observable system is a system that can always define some neighborhood $U$ around every $x$ such that there are no indistinguishable state pairs in that neighborhood. However, weak observability says nothing about how small the neighborhood $U$ should be. The concept of local weak observability is a more strict definition. A system is locally
weakly observable at \( x_0 \) if there exists a neighborhood \( U \) of \( x_0 \) such that for any neighborhood \( V \) of \( x_0 \) contained in \( U \), there is no indistinguishable state from \( x_0 \) in \( V \), when considering time intervals for which trajectories remain in \( V \). This essentially means that a locally weakly observable system can distinguish every state from its neighbor without “going too far” [13].

The matrix \( O' \) is analogous to the observability matrix for linear systems. In fact, it can be shown that the observability matrix using the Lie derivative approach becomes the same as Eq. (3.14), if the system is linear. Consider a 2 state, LTI system with no input and a single output:

\[
\begin{bmatrix}
\dot{x}_1 \\
\dot{x}_2
\end{bmatrix} =
\begin{bmatrix}
a_{1,1} & a_{1,2} \\
a_{2,1} & a_{2,2}
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2
\end{bmatrix} = Ax = f(x)
\]

\[
y = \begin{bmatrix} c_1 & c_2 \end{bmatrix} \begin{bmatrix} x_1 \\
x_2 \end{bmatrix} = Cx = g(x)
\]

The Lie derivatives become:

\[
\begin{bmatrix}
L^0_f \left( g \right) \\
L^1_f \left( g \right)
\end{bmatrix} =
\begin{bmatrix}
g(x) \\
\frac{\partial g}{\partial x_1} f_1(x) + \frac{\partial g}{\partial x_2} f_2(x)
\end{bmatrix} =
\begin{bmatrix}
c_1 x_1 + c_2 x_2 \\
c_1 (a_{1,1} x_1 + a_{1,2} x_2) + c_2 (a_{2,1} x_1 + a_{2,2} x_2)
\end{bmatrix}
\]

Finally, taking the Jacobian gives:

\[
O' =
\begin{bmatrix}
c_1 & c_2 \\
c_1 a_{1,1} + c_2 a_{2,1} & c_1 a_{1,2} + c_2 a_{2,2}
\end{bmatrix} =
\begin{bmatrix}
C \\
CA
\end{bmatrix}
\]

which is the same result as Eq. (3.14).
3.2.3 Observability of the single particle model

The focus of this work is to design state and parameter observers for the single particle model outlined in Chapter 2; however, it should first be determined if the model is observable. In fact, since the only output measurement available in an automotive application is the cell voltage (essentially the difference between each electrode potential), it is intuitive that it may be difficult to observe the states of both electrodes. For example, if a cell voltage of 4 V is measured, absent any other information this could be the result of 4.2 V from the positive electrode and 0.2 V from the negative, or equally likely, 4.3 V from the positive and 0.3 V from the negative. It is not clear whether or not a unique state vector for both electrodes can be found, and testing the model observability is an essential step.

In the single particle model, the input is current, output is cell voltage, and the states are related to the lithium concentrations in the solid or liquid phase. Lithium concentrations in the electrodes are analogous to SOC (this is elaborated on in Chapter 4), so analyzing observability for the model states gives insight into whether it is possible to estimate SOC. For SOH estimation, additional model parameters that change with age may be included as states and estimated along with the original states, in which case the observability of the new state vector must be determined. For now, the following analysis will just consider the original model as outlined in Chapter 2, where the states are related to concentrations.

The nonlinear behavior in the output voltage necessitates taking the Lie derivative approach to prove local observability. Due to the relatively large number of states and the complexity of the output equation, the resulting observability matrix is too lengthy to show here. However, by making some simplifications, it is possible to show the process of calculating the Lie derivatives, and still draw some conclusions about the observability properties of the model. The first simplification is to consider only
a first order Padé approximation of the solid diffusion equation (Eq. (2.48)). Second, the charge transfer overpotentials (Eq. (2.107)) and liquid phase overpotential (Eq. (2.126)) can both be treated as Ohmic. Finally, the OCV term for electrode \(i\), as a function of concentration, can be defined by a cubic spline, \(U_i = s_{0,i}c_i^3 + s_{1,i}c_i^2 + s_{2,i}c_i + s_{3,i}\), where \(s_{j,i}\) are the spline coefficients for a given segment. Note that the spline segments are matched so that the entire curve is continuously differentiable up to the second derivative. The model for a single material electrode pair then becomes:

\[
\begin{bmatrix}
\dot{x}_p \\
\dot{x}_n
\end{bmatrix} = f(x, u) = \begin{bmatrix}
\frac{R_p}{3FA L_p} u \\
-\frac{R_n}{3FA L_n} u
\end{bmatrix}
\]

where the bulk concentrations are proportional to the states:

\[
\begin{bmatrix}
\bar{c}_p \\
\bar{c}_n
\end{bmatrix} = \begin{bmatrix}
a_{0,p} \\
a_{0,n}
\end{bmatrix} x.
\]

The Lie derivatives are:

\[
L_{0j}^0(g) = g(x)
\]

\[
= s_{0,p}a_{0,p}^3x_1^3 + s_{1,p}a_{0,p}^2x_1^2 + s_{2,p}a_{0,p}x_1 + s_{3,p} - s_{0,n}a_{0,n}x_2^3 - s_{1,n}a_{0,n}x_2^2 - s_{2,n}a_{0,n}x_2 - s_{3,n} - R_{\text{ohmic}} u
\]

\[
L_{1j}^1(g) = \frac{\partial g}{\partial x_1} f_1(x) + \frac{\partial g}{\partial x_2} f_2(x) + \frac{\partial g}{\partial u} \dot{u}
\]

\[
= s_{0,p}a_{0,p}^3x_1^2R_p u + 2s_{1,p}a_{0,p}^2x_1R_p u + s_{2,p}a_{0,p}R_p u + \frac{s_{0,p}a_{0,p}^3x_2^3R_n u}{FA L_p} + \frac{2s_{1,p}a_{0,p}^2x_2R_n u}{3FA L_p} + \frac{s_{2,p}a_{0,p}R_n u}{3FA L_p} + \frac{s_{0,n}a_{0,n}x_2^3R_n u}{FA L_n} + \frac{2s_{1,n}a_{0,n}x_2R_n u}{3FA L_n} + \frac{s_{2,n}a_{0,n}R_n u}{3FA L_n} - R_{\text{ohmic}} u
\]
Taking the Jacobian with respect to $x$ gives the observability matrix:

$$
\mathcal{O}' = \begin{bmatrix}
3s_{0,p}a_0^3p^2x_1^2 + 2s_{1,p}a_0^2p^2x_1 + s_{2,p}a_{0,p} & -3s_{0,n}a_{0,n}^3x_2^2 - 2s_{1,n}a_{0,n}^2x_2 - s_{2,n}a_{0,n} \\
\frac{2s_{0,p}a_0^3p^2x_1R_p}{FAL_p\epsilon_p} + \frac{2s_{1,p}a_0^2p^2R_p}{3FAL_p\epsilon_p} & \frac{2s_{0,n}a_{0,n}^3x_2R_n}{FAL_n\epsilon_n} + \frac{2s_{1,n}a_{0,n}^2R_n}{3FAL_n\epsilon_n}
\end{bmatrix}
$$

(3.30)

It is easy to see that this matrix has full rank, other than for some fluke occurrence when the specific values of the parameters, states, and input align such that the two columns or two rows become equal. Therefore, the states of this simplified system are indeed locally observable from the cell voltage measurement. It is interesting to note that the nonlinearity in the output and the non-zero current are the only reasons why this is the case. If the OCVs were linear ($s_{0,p} = s_{1,p} = s_{0,n} = s_{1,n} = 0$), then the observability matrix becomes:

$$
\mathcal{O}' = \begin{bmatrix}
s_{2,p}a_{0,p} & -s_{2,n}a_{0,n} \\
0 & 0
\end{bmatrix}
$$

(3.31)

and loses rank. Alternatively, if the input current is zero, the observability matrix becomes:

$$
\mathcal{O}' = \begin{bmatrix}
3s_{0,p}a_0^3p^2x_1^2 + 2s_{1,p}a_0^2p^2x_1 + s_{2,p}a_{0,p} & -3s_{0,n}a_{0,n}^3x_2^2 - 2s_{1,n}a_{0,n}^2x_2 - s_{2,n}a_{0,n} \\
0 & 0
\end{bmatrix}
$$

(3.32)

and loses rank. This system would be classified as locally observable, but not linearly observable or zero input observable, according to the definitions in [53].

A similar analysis can be done on the original model with the 3rd order Padé approximation and non-Ohmic charge transfer overpotential, as shown in Appendix A.1, also resulting in an observability matrix with full rank. However, the 3rd order approximation results in an observability matrix that does not lose rank even for
zero current, assuming that the cell is not at steady-state (i.e. the cell is relaxing following a previous current pulse). There are additional state dynamics for the 3rd order approximation when the current is zero that help to avoid the loss of rank. These dynamics when $I = 0$ are not present for the 1st order approximation (since when $I = 0$, $\dot{x} = 0$). That said, even the 3rd order approximation model becomes unobservable if the current is zero and the cell has reached steady state. Steady state implies that $x_2 = x_3 = x_5 = x_6 = 0$, in which case the 3rd order and 1st order models essentially become identical. Returning to the example posed at the beginning of this section, if a cell reads 4 V and is at steady state, there should be no way to distinguish between electrode voltages of say, 4.2/0.2 V and 4.3/0.3V. The observability analysis confirms this intuition. However, the nonlinearities in the output and relaxation dynamics of the higher order Padé approximation allow the states to be observed for cases when the cell is not at steady state.

While the states of the complete model are theoretically observable, practically speaking this does not mean that the model structure is suitable for an observer. It is possible that numerically the observability matrix is very close to losing rank. The condition number of the observability matrix can be used as a measure of “degree of observability”. The condition number of a matrix $A$ can be calculated as:

$$\kappa(A) = \|A^{-1}\| \cdot \|A\|$$  \hspace{1cm} (3.33)

where $\| \cdot \|$ indicates the 2-norm. Conceptually, the condition number is a measure of how measurement errors (either due to noise, limited resolution, or model error) contribute to errors in the estimated state vector [27]. A high condition number indicates that measurement errors are greatly amplified and are projected into large state estimate errors. For example, recall the discrete linear system observer problem,
where a series of measurement samples are used to reconstruct the state vector:

\[
\begin{bmatrix}
 y_0 - Du_0 \\
 y_1 - Du_1 - CB_d u_0 \\
 y_2 - Du_2 - CB_d u_1 - CA_d B_d u_0
\end{bmatrix} =
\begin{bmatrix}
 C \\
 CA_d \\
 CA_d^2
\end{bmatrix} x_0
\]

\[ z = M x_0 \]

Invertability of the observability matrix, \( M \), implies a unique solution for \( x_0 \). However, now consider some error in the measurement, defined as \( \Delta z \). This measurement error will produce an error in the calculated \( x_0 \) once the matrix inversion is performed:

\[ M^{-1}(z + \Delta z) = x_0 + \Delta x_0 \]

Essentially, the condition number measures the ratio of the relative error in \( x_0 \) to the relative error in \( z \):

\[ \kappa(M) = \sup \left( \frac{\| \Delta x_0 \|}{\| x_0 \|} / \frac{\| \Delta z \|}{\| z \|} \right) \]

Therefore, a high condition number indicates an observability matrix that greatly amplifies measurement errors, leading to an inaccurate reconstruction of the state vector. The condition number of the observability matrix resulting from the single particle model is analyzed in Chapter 4. In short, the model that includes both electrodes results in a poorly conditioned observability matrix, motivating various approaches to improve the degree of observability.

### 3.3 Observer Design

There are many different techniques for performing state estimation. Some techniques like fuzzy logic or neural network estimators [41] do not require a model, and simply
establish some correlation between inputs, outputs, and states. For SOC/SOH estimation, the interest in this work is to estimate states using a model-based approach with an observer. An observer is an algorithm that uses a model of a system and the measured inputs and outputs to estimate the internal states, which may not be directly measurable. Even if an internal state is measurable, an observer could be used to make a redundant estimate of that quantity in order to filter the signal or detect a sensor fault. Many observers use a prediction-correction (or prediction-update) scheme at each time step, as shown in Fig. 3.1. The model is first used to predict the states and output based on the known input. Then those predicted states are corrected based on the difference between the predicted output and the measured output (called the output residual). In general, if there is a large difference between the predicted and measured output, then the predicted states were not very accurate and a large correction is in order. Model-based observers mainly differ in how that state correction is calculated and applied. For example, the extended Kalman filter calculates the state correction based on a linearized version of the model equations. The particle filter and unscented Kalman filter take an approach where many possible versions of the system states are generated, and each version is weighted based on how likely it is to represent the true system. The state correction is then scheduled based on these weights.

The extended Kalman filter and particle filter are outlined in detail in the following sections, since these were the algorithms used for SOC/SOH estimation in this work; however, other common nonlinear observers will be described briefly as well.

For each algorithm, the model is assumed to be implemented in discrete time. For example, the particle sub-model is implemented as:

\[
x_k = f_{k-1}(x_{k-1}, u_{k-1}, w_{k-1}) = (I_d + \Delta t A)x_{k-1} + \Delta t B_{d}u_{k-1} + w_{k-1}
= A_d x_{k-1} + B_{d,k-1} u_{k-1} + w_{k-1}
\]  

(3.37)
Figure 3.1: Typical structure of a model-based observer, involving prediction, output residual generation, and correction.

\[ y_k = g_k(x_k, u_k, v_k) = U(x_k) + \eta(x_k, \beta_k u_k) + R_c \beta_k u_k + v_k \]  \hspace{1cm} (3.38)

where \( x \) is the state vector, \( y \) is the output voltage, \( u \) is the input, \( w \) is the process noise, \( v \) is the sensor noise, \( A_d \) and \( B_d \) are the state dynamics matrices in discrete time (\( B_d \) is allowed to be time varying to accommodate the composite electrode), \( k \) is the current time step, and \( \Delta t \) is the sampling interval (fixed at 0.1 s). \( Q \) and \( R \) are defined as the covariance matrices of the process noise and sensor noise, respectively.

### 3.3.1 The extended Kalman filter

The Kalman filter (KF) is a well-known algorithm for state estimation due to its ease of implementation and optimality under certain conditions. The requirements for optimality are that the process and sensor noise be white zero-mean Gaussian noise with known covariance matrices and that the system have known linear dynamics and output. The samples in white noise are characterized by having a constant power spectrum and being independent (and therefore uncorrelated). A zero-mean Gaussian distribution means that \( p(w) \sim \mathcal{N}(0, Q) \) and \( p(v) \sim \mathcal{N}(0, R) \). If these requirements are met, then the Kalman filter finds the minimum mean squared error estimate of the state vector, \( \hat{x}_k \) (with respect to the true state \( x_k \)), using the available
input and output data up to and including the current time step $k$. In other words, the Kalman filter problem is to solve:

$$
\hat{x}_k = \arg\min_{\hat{x}} \mathbb{E}[(x_k - \hat{x})^T(x_k - \hat{x})|u_0, u_1, \ldots, u_k, y_0, y_1, \ldots, y_k] \quad (3.39)
$$

where $\mathbb{E}$ is the expected value operator used in probability and statistics. The solution can be implemented as a recursive algorithm, where at each time step, estimates of the state vector $\hat{x}_k$ as well as the covariance matrix of the state estimate error, $P_k = \mathbb{E}[(x_k - \hat{x}_k)^T(x_k - \hat{x}_k)]$ are obtained. The estimate error covariance can be used as a measure of the estimate uncertainty, in order to establish error bars on the estimate. The recursive nature of the algorithm means that the $\hat{x}_k$ and $P_k$ estimates at the current time step are only dependent on the estimates and measurements at the previous and current time steps. The algorithm minimizes the estimate error over the entire available measurement data set without actually needing to keep track of the entire set of estimates and measurements. This makes the algorithm computationally efficient since it allows the estimates and measurements to be discarded as time progresses, if desired.

In practice, the process and sensor noise are never perfectly white, zero-mean, or Gaussian, and the noise covariances may not be known. Additionally no real system is linear and no model is perfect; however despite this, the Kalman filter is robust enough to still perform well in many cases.

At each time step, the Kalman filter algorithm performs a prediction step and a correction step. In the prediction step, the state vector, output, and estimate covariance at the current time step are predicted using only knowledge of their previous values and the input. These are known as a priori estimates (indicating that they occur before considering the measurement), and are denoted by a superscript $\sim$. In the correction step, the predicted states and covariance are corrected by considering
how different the predicted output is from the measured output. These are known as *a posteriori* estimates (indicating that they occur after considering the measurement), and are denoted by a superscript +. The magnitude of the correction is determined by the magnitude of the output residual, as well as the gain applied to that difference.

For a generic linear system in discrete time, \( x_k = A_d x_{k-1} + B_d u_{k-1} + w_{k-1}, y_k = C x_k + D u_k \), the Kalman filter algorithm is as follows:

1. Initialize the filter at \( k = 0 \):

   \[
   \hat{x}_0^+ = \mathbb{E}[x_0] \quad (3.40)
   \]

   \[
   P_0^+ = \mathbb{E}[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T] \quad (3.41)
   \]

   Since the true state \( x_k \) is not known, a best guess can be used to initialize \( \hat{x}_0^+ \) and the expected estimate error can be set to zero. The whole idea of the filter is that the estimate will converge to the true value even if it is initialized incorrectly.

2. At time step \( k \), the predicted state vector and output are advanced in time according to the model dynamics:

   \[
   \hat{x}^-_k = A_d \hat{x}^+_k + B_d u_{k-1} \quad (3.42)
   \]

   \[
   \hat{y}^-_k = C \hat{x}^-_k + D u_k \quad (3.43)
   \]

3. The predicted state estimate covariance is advanced in time according to:

   \[
   P^-_k = A P^+_k A^T + Q \quad (3.44)
   \]

4. The predicted state estimate is then corrected using feedback from the mea-
surement, $y_k$.

\[
\hat{x}_k^+ = \hat{x}_k^- + K_k(y_k - \hat{y}_k^-) \tag{3.45}
\]

The Kalman gain, $K$, determines how much of an effect the output residual (also called the innovation) has on correcting the states, and is defined as.

\[
K_k = P_k^- C^T (C P_k^- C^T + R)^{-1} \tag{3.46}
\]

5. Finally, the corrected estimate covariance is updated:

\[
P_k^+ = (I_d - K_k C) P_k^- \tag{3.47}
\]

The corrected covariance is always made smaller than the previous predicted covariance, since a new measurement is being considered.

For nonlinear systems, the basic Kalman filter algorithm only needs to be slightly modified. Essentially, linear $A$ and $C$ matrices can be approximated by linearizing the nonlinear dynamics and output at each time step, i.e.

\[
A_k \approx \frac{\partial f}{\partial x} \bigg|_{\hat{x}_k^+} \text{ and } C_k \approx \frac{\partial g}{\partial x} \bigg|_{\hat{x}_k^-}.
\]

The linearizations are then used to calculate the Kalman gain and propagate the estimated error covariance. This leads to the well-known extended Kalman filter (EKF). The EKF algorithm equations are outlined in Table 3.1 for a generic nonlinear system in discrete time. The approximations required in the EKF result in a state estimate that is no longer optimal, even for perfectly white zero-mean Guassian noise and a perfect model. However, it works well for many battery estimation applications [71–73]. For certain systems that are highly nonlinear or exhibit discontinuous behavior, the linearization approximation may be poor. In this case, other nonlinear extensions to the Kalman filter such as the unscented Kalman filter may be used, which do not rely on linearization.
Table 3.1: Extended Kalman filter algorithm for state estimation of a generic nonlinear system with assumed process and sensor noise.

<table>
<thead>
<tr>
<th>Step</th>
<th>Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generic Nonlinear System</strong></td>
<td>$x_k = f(x_{k-1}, u_{k-1}) + w_{k-1}$</td>
</tr>
<tr>
<td></td>
<td>$y_k = g(x_k, u_k) + v_k$</td>
</tr>
<tr>
<td><strong>Initialization</strong></td>
<td>$\hat{x}_0^+ = \mathbb{E}(x_0)$</td>
</tr>
<tr>
<td></td>
<td>$P_0^+ = \mathbb{E}[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T]$</td>
</tr>
<tr>
<td></td>
<td>$\hat{x}_0^- = f(\hat{x}<em>0^+; u</em>{k-1})$</td>
</tr>
<tr>
<td></td>
<td>$\hat{y}_0^- = g(\hat{x}_0^-, u_k)$</td>
</tr>
<tr>
<td><strong>Prediction Step</strong></td>
<td>$\hat{x}<em>k^- = A_k P</em>{k-1}^- A_k^T + Q$</td>
</tr>
<tr>
<td></td>
<td>$A_k = \frac{\partial f}{\partial x} \bigg</td>
</tr>
<tr>
<td></td>
<td>$\hat{x}_k^+ = \hat{x}_k^- + K_k (y_k - \hat{y}_k^-)$</td>
</tr>
<tr>
<td></td>
<td>$K_k = P_k^- C_k^T (C_k P_k^- C_k^T + R)^{-1}$</td>
</tr>
<tr>
<td><strong>Correction Step</strong></td>
<td>$C_k = \frac{\partial g}{\partial x} \bigg</td>
</tr>
<tr>
<td></td>
<td>$P_k^+ = (I_d - K_k C_k) P_k^-$</td>
</tr>
</tbody>
</table>

The noise covariances $Q$ and $R$ have a large effect on the performance of the filter. The sensor noise is typically easy to compute, simply by observing the sensor behavior over some period of time. However, the process noise is often not possible to quantify, since there is not necessarily a direct measurement of the states. Therefore, $Q$, and sometimes $R$ as well, are often used as tuning parameters to achieve a desired filter performance. It is evident that increasing $R$ will decrease the Kalman gain $K$ and provide a smaller state correction for a given output residual. A high $R$ is telling the filter to not trust the measurement and avoid making large state corrections based on the uncertain measurement. For such a case, the state convergence from an initial condition error will be slow. On the other hand, increasing $Q$ will increase the estimate covariance and ultimately increase the Kalman gain. A high $Q$ is telling the filter that the predicted states using just the model are very uncertain, so the measurement will be heavily relied on to make large corrections to the states. In this
Figure 3.2: Example of the effect tuning the process noise covariance, $Q$, has on the state estimate.

case convergence from an initial condition error will be fast, but the estimate may be very noisy if $Q$ is tuned too high.

An example of how tuning $Q$ and $R$ affects the state estimate is shown in Figs. 3.2 - 3.3, for the single particle model. In this example, a perfect model is used to generate noisy pseudo-data to use as a voltage “measurement”.

Intrinsically, the presence of unknown model error is something that an observer cannot hope to overcome; the estimate accuracy is ultimately limited by the model accuracy. However, the filter performance in the presence of model error can be improved under certain conditions by tuning $Q$ and $R$. Model error can be thought
Figure 3.3: Example of the effect tuning the sensor noise covariance, $R$, has on the state estimate.

of as an unknown addition to the output residual. A large model error will appear to the filter as a large output residual that will tend to trigger a large state correction. To avoid these large state corrections, $Q$ can be decreased and/or $R$ can be increased so as to effectively decrease the Kalman gain. The result will be a state estimate that has slow convergence but that does not overreact and make large corrections in the presence of model error.
3.3.2 The particle filter

To discuss the particle filter, it is helpful to step back and review general Bayesian estimation from a probabilistic perspective. For some generic nonlinear discrete system with process and sensor noise:

\[ x_k = f(x_{k-1}, u_{k-1}) + w_{k-1} \]  \hspace{1cm} (3.48)

\[ y_k = g(x_k, u_k) + v_k \]  \hspace{1cm} (3.49)

the Bayesian estimation problem is to construct the posterior probability density function (PDF) of the state at time \( k \), based on all available information, including the available measurements, i.e. \( p(x_k|y_1,\ldots,k) \) [6,81]. A state estimate can then be obtained from this PDF, that is optimal with respect to any desired criteria. The posterior PDF also contains information about the uncertainty associated with the estimate. In the recursive prediction-correction structure, this amounts to first constructing the prior PDF at time \( k \) (prediction step) using the Chapman-Kolmogorov equation:

\[ p(x_k|y_1,\ldots,k-1) = \int p(x_k|x_{k-1})p(x_{k-1}|y_1,\ldots,k-1)dx_{k-1} \]  \hspace{1cm} (3.50)

where \( p(x_k|x_{k-1}) \) is known from the model of the state dynamics and process noise (Eq. (3.48)), and \( p(x_{k-1}|y_1,\ldots,k-1) \) is simply the posterior PDF at the previous time step. After introducing the measurement at time \( k \), the posterior PDF (correction step) is calculated using Bayes’ rule:

\[ p(x_k|y_1,\ldots,k) = \frac{p(y_k|x_k)p(x_k|y_1,\ldots,k-1)}{\int p(y_k|x_k)p(x_k|y_1,\ldots,k-1)dx_k} \]  \hspace{1cm} (3.51)
where \( p(y_k|x_k) \) is known from the model of the output and sensor noise (Eq. (3.49)). Unfortunately, finding an analytical solution to this problem is not possible, except in rare cases. One such case is under the assumption of linear dynamics and output, and white zero-mean Gaussian noise, which leads to the Kalman filter. However in most other cases, it is only possible to approximate the posterior PDF. This is essentially what the extended Kalman filter is doing, by linearizing the state and output dynamics. The particle filter takes a different approach, but the goal is still to approximate the posterior PDF.

Rather than estimate a single state vector at each time step, such as in the EKF, the particle filter estimates a distribution of state vectors at each time step, where each potential state vector and corresponding output is referred to as a particle. Therefore, each particle is a possible representation of the true system. The set of particles can then be assigned weights based on their likelihood to represent the true state. The set of weighted particles is an approximation of the posterior PDF. An estimate of this posterior PDF allows for an estimate of the states themselves (typically from the mean of the estimated PDF) as well as the uncertainty associated with the state estimate (for example, by analyzing the variance of the estimated PDF). The particle filter approach is essentially like trying to hit a bull’s-eye on a dart board by randomly throwing a group of darts at the board repeatedly, where between each set of throws, the trajectories that led to bad misses are made less likely and the trajectories that led to near hits are made more likely. For the next set of throws, the grouping will still be random, but it will be more tightly packed around the bull’s-eye. The particle filter is a type of Monte Carlo algorithm, in that it involves repeated random sampling. It is also referred to as a sequential Monte Carlo approach or a bootstrap filter.

To outline the algorithm equations, the process and sensor noise are assumed to be white zero-mean Gaussian, although this need not be the case. The algorithm consists
of 3 steps for each time step iteration: prediction, correction, and re-sampling.

1. To initialize the filter, the initial posterior PDF, \( p(x_0|y_0) \), is assumed to be Gaussian with mean according to the best guess of the initial state vector and variance that is tuned to give a large spread of particles. Particles are then randomly sampled from this initial PDF. The word “particle” in this section should not be confused with an electrode particle.

2. For each particle, \( i \), the states, \( \hat{x}_i \), and output, \( \hat{y}_i \), at the next time step are predicted from the model:

\[
\hat{x}_{i,k}^- = f(\hat{x}_{i,k-1}^+, u_{k-1}, w_{i,k-1})
\]

\[
\hat{y}_{i,k} = g_k(\hat{x}_{i,k}^-, u_k)
\]

where \( w_{i,k-1} \) is a sample drawn from the PDF of the process noise, assumed here to be zero-mean Gaussian with covariance matrix \( Q \).

3. Now the measurement at time step \( k \) is introduced as \( y_k \). Since there is noise in the measurement, in theory, any of the predicted states and output could represent reality, even if the measured \( y_k \) is not close to the predicted \( \hat{y}_{i,k} \). Therefore, the goal of the correction step is to decide which particles are more likely to represent reality, without knowledge of the true output but with a noisy measurement of the output. This translates to evaluating, for each particle, the conditional probability of seeing the measurement, \( y_k \), given the predicted states and output, \( \hat{x}_{i,k}^- \) and \( \hat{y}_{i,k} \), are the true values, i.e. \( p(y_k|\hat{x}_{i,k}^-) \). The conditional probability is used to assign a normalized weight to each particle, \( q_i \), indicating
the likelihood that it represents reality:

\[ q_i = \frac{p(y_k | \hat{x}_{i,k}^-)}{\sum_{i=1}^{N_p} p(y_k | \hat{x}_{i,k}^-)} \]  
(3.54)

For the assumed Gaussian distribution of sensor noise with zero mean and variance \( R \), the conditional probability for each particle evaluates to [21]:

\[ p(y_k | \hat{x}_{i,k}^-) = \frac{1}{\sqrt{2\pi R}} \exp\left(\frac{-(y_k - \tilde{y}_{i,k})^2}{2R}\right) \]  
(3.55)

The weighting process defines a new discrete probability mass function over the particles, \( \{\hat{x}_{i,k}^- : i = 1, ..., N_p\} \), where element \( i \) has a probability mass of \( q_i \).

4. Now new particles are generated by resampling \( N_p \) times from the weighted discrete probability mass function. Thus, corrected state estimates, \( \hat{x}_{i,k}^+ \) are generated so that for any \( j \), the probability of \( \hat{x}_{j,k}^+ \) being equal to \( \hat{x}_{i,k}^- \) is \( q_i \) (i.e. \( Pr\{\hat{x}_{j,k}^+ = \hat{x}_{i,k}^-\} = q_i \)). The distribution of the corrected state estimates \( \hat{x}_{i,k}^+ \) is approximately equal to the posterior PDF, \( p(x_k | y_1, ..., y_k) \). Finally, a single state estimate at time \( k \) is obtained by taking the mean of all \( \hat{x}_{i,k}^+ \).

The particle filter algorithm can be visualized in Fig. 3.4. The number of particles is an important parameter that offers a tradeoff between accuracy and computational effort. More particles allows the weighted discrete probability mass function to better approximate the posterior PDF; however, each particle requires its own prediction and correction calculations, so adding a particle increases the computational effort a similar amount to say adding another model state in an EKF algorithm. The particle filter shows good performance, especially for highly nonlinear systems which make the linearization approach used by the EKF undesirable.
1. Randomly distribute \( N \) particles around the initial guess of the state, according to some distribution.

2. Propagate the state and output forward in time for each particle.

3. For each particle, calculate \( P(y_k | \hat{y}_{i,k}) \).
   For example, the observed measurement is not very likely for particle \( i \) shown, so particle \( i \) is not likely to be “true”.

4. Form a weighted distribution based on the probability of each particle to be “true”.

Figure 3.4: Visualization of the particle filter algorithm.
### 3.3.3 Other nonlinear observers

Other common model-based nonlinear observers include the Luenberger observer and unscented Kalman filter.

**Luenberger observer**

The Luenberger observer has a very similar structure to the Kalman filter, except that the gain applied to the output residual is constant [40]. The predicted state is corrected according to:

\[
\hat{x}_k^+ = \hat{x}_k^- + L(y_k - y_k^-) \tag{3.56}
\]

where \( L \) is a constant matrix that is designed to give the desired observer performance. The dynamics of the estimate error, \( e \), for a discrete linear system become:

\[
\dot{e}_{k+1} = (A_d - A_dLC)e_k \tag{3.57}
\]

Specifically, the eigenvalues of \((A_d - A_dLC)\) that dictate the estimate error trajectory over time can be selected by an appropriate selection of the elements in \( L \). Clearly, \( L \) should be selected so that the estimate error decays to zero (eigenvalues < 0), and does so reasonably fast.

To extend the Luenberger observer to nonlinear systems, one approach could be to approximate the system as linear around some operating point in order to select \( L \); however, guaranteeing the estimate error stability using this approach is not always possible. Stability can be guaranteed using Lyapunov-based methods for certain classes of nonlinear systems [74].

113
Unscented Kalman filter

The unscented Kalman filter (UKF) is another extension to the Kalman filter for nonlinear systems, that takes a different approach than the EKF. Similar to the particle filter, the UKF attempts to approximate the posterior PDF of the state estimate by a discrete set of points (called sigma points in this case), chosen so that their sample mean and sample covariance approximate the true mean and covariance [48]. The major difference between the particle filter and the UKF is that in the UKF, the sigma points are not drawn at random like the particles in the particle filter, but rather according to a deterministic algorithm. For a state vector with \( n \) states, the estimate mean, \( \hat{x} \), and covariance, \( P_{xx} \), are approximated by \( 2n + 1 \) weighted sigma points. The sigma points, \( \chi \), and weights, \( W \), are given by:

\[
\begin{align*}
\chi_0 &= \hat{x} \\
\chi_i &= \hat{x} + \left( \sqrt{(n + \kappa)P_{xx}} \right)_i \\
\chi_i+n &= \hat{x} - \left( \sqrt{(n + \kappa)P_{xx}} \right)_i
\end{align*}
\]

\[
\begin{align*}
W_0 &= \frac{\kappa}{n + \kappa} \\
W_i &= \frac{1}{2(n + \kappa)} \\
W_{i+n} &= \frac{1}{2(n + \kappa)}
\end{align*}
\]

where \( i = 1, \ldots, n \), the term \( \left( \sqrt{(n + \kappa)P_{xx}} \right)_i \) is the \( i \)th row or column of the matrix square root of \( (n + \kappa)P_{xx} \), and \( \kappa \) is used as a tuning parameter. For a Gaussian process noise, selecting \( \kappa \) so that \( n + \kappa = 3 \) is a good starting choice. The procedure for the UKF is shown in Table 3.2 [48,57]. Note that unlike the EKF, no linearization of the system is required. So while additional calculation is required for the \( 2n + 1 \) sigma points (compared to the EKF), the elimination of the linearization process results in a computational complexity that can be similar between the UKF and EKF [57].
Table 3.2: Unscented Kalman filter algorithm for a generic nonlinear system with assumed process and sensor noise.

<table>
<thead>
<tr>
<th>Step</th>
<th>Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generic Nonlinear System</strong></td>
<td>( x_k = f(x_{k-1}, u_{k-1}) + w_{k-1} ) &lt;br&gt; ( y_k = g(x_k, u_k) + v_k )</td>
</tr>
<tr>
<td><strong>Initialization at</strong> ( k = 0 )</td>
<td>( \hat{x}_0^+ = \mathbb{E}(x_0) ) &lt;br&gt; ( P_0^+ = \mathbb{E}[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T] )</td>
</tr>
<tr>
<td><strong>Generate</strong> ( j ) sigma points and weights, where ( j = 1, \ldots, 2n + 1 )</td>
<td>( \chi_0 = \hat{x}<em>{k-1}^+ ) &lt;br&gt; ( \chi_i = \hat{x}</em>{k-1}^+ + \left( \sqrt{(n + \kappa)P_{k-1}^+} \right)<em>i ) &lt;br&gt; ( \chi</em>{i+n} = \hat{x}<em>{k-1}^+ - \left( \sqrt{(n + \kappa)P</em>{k-1}^+} \right)<em>i ) &lt;br&gt; ( W_0 = \frac{\kappa}{n+\kappa} ) &lt;br&gt; ( W_i = \frac{1}{2(n+\kappa)} ) &lt;br&gt; ( W</em>{i+n} = \frac{1}{2(n+\kappa)} )</td>
</tr>
<tr>
<td><strong>Prediction step for each sigma point</strong></td>
<td>( \chi_{j,k} = f(\chi_{j,k-1}, u_{k-1}) ) &lt;br&gt; ( Y_{j,k} = g(\chi_{j,k}, u_k) )</td>
</tr>
<tr>
<td><strong>Compute predicted state and output mean and covariance</strong></td>
<td>( \hat{x}<em>k = \sum</em>{j=0}^{2n} W_j \chi_{j,k} ) &lt;br&gt; ( \hat{y}<em>k = \sum</em>{j=0}^{2n} W_j Y_{j,k} ) &lt;br&gt; ( P_k^- = \sum_{j=0}^{2n} W_j (\chi_{j,k} - \hat{x}<em>k^-)(\chi</em>{j,k} - \hat{x}_k^-)^T + Q )</td>
</tr>
<tr>
<td><strong>Compute innovation covariance and cross covariance</strong></td>
<td>( P_{yk} = R + \sum_{j=0}^{2n} W_j (Y_{j,k} - \hat{y}<em>k)(Y</em>{j,k} - \hat{y}<em>k)^T ) &lt;br&gt; ( P</em>{ykx} = \sum_{j=0}^{2n} W_j (\chi_{j,k} - \hat{x}<em>k^-)(Y</em>{j,k} - \hat{y}_k)^T )</td>
</tr>
<tr>
<td><strong>Correction Step</strong></td>
<td>( K_k = P_{ykx} P_{yk}^{-1} ) &lt;br&gt; ( \hat{x}_k^+ = \hat{x}_k^- + K_k (y_k - \hat{y}<em>k) ) &lt;br&gt; ( P_k^+ = P_k^- - K_k P</em>{yk} K_k^T )</td>
</tr>
</tbody>
</table>

3.4 Methods for Smoothing

The goal of the Kalman filter is to minimize the mean squared estimate error using all available measurements up to and including the current time step. The state estimate improves with each additional measurement. However, it is intuitive that the estimate at time \( k \) could be improved further still by waiting until more measurements become available after time \( k \). This method is known as smoothing and is a type of batch estimation. The obvious drawback is that the estimate at time \( k \) cannot be made in
real time, since it is now dependent on past, present, and future measurements.

There are three basic types of smoothing techniques: fixed point, fixed lag, and fixed interval smoothing. For fixed point smoothing, the estimate of interest is at a fixed time step $k$. As more measurements come in after time $k$, the estimate at time $k$ continues to improve. An example might be if a satellite takes a picture at time $k$, but needs estimates of its position and velocity in order to process the picture. As more measurements come in after time $k$, the position and velocity estimates continue to improve and the picture becomes clearer [88]. For fixed lag smoothing, the estimate of interest is always at a fixed amount of time in the past. In other words, the goal is to estimate the state at time $k - N$, where $k$ continually changes with time, but $N$ is fixed. Using the same satellite example, a fixed lag smoother might be used if the picture must be transmitted to Earth $N$ time steps after it is taken. The processing is allowed to improve the picture with additional measurements for $N$ time steps, but after that it must be transmitted. The delay in receiving each picture is fixed [88]. Finally, for fixed interval smoothing, some interval of measurements $[y_1, \ldots, y_N]$ is available, and the best estimate at each time step is desired. The estimate at each time step takes into account all of the available measurements over the entire interval. In this way, the estimate at $k = 1$ is greatly improved compared to the traditional filter, since it considers all $N$ measurements rather than just one. On the other hand, the estimate at $k = N$ does not improve compared to a filter, since in the filter, all $N$ measurements were already being considered. Fixed interval smoothing is most applicable to the SOH estimation problem. For example, current and voltage data can be collected over the course of a drive cycle, parameters can then be estimated for that fixed interval of data, and ultimately updated in the model once the estimates are obtained. However, this requires that the input and output data for the drive cycle be stored onboard the vehicle until the state estimate is performed.

In the fixed interval Kalman smoother, the traditional Kalman filter is essentially
run once forward in time and then again backward in time. This means that two state
estimates, a forward and backward estimate, are obtained at each time step \( k \). The
forward estimate takes into account measurements \([y_1, \ldots, y_k]\), while the backward
estimate takes into account measurements \([y_N, \ldots, y_k]\). The forward and backward
estimates, \( \hat{x}_{f,k} \) and \( \hat{x}_{b,k} \), are ultimately combined in an optimal way to arrive at the
final smoothed estimate. For a linear system, the model for the forward filter is
unchanged:

\[
x_{f,k} = Ax_{f,k-1} + Bu_{k-1} + w_{k-1} \tag{3.59}
\]

\[
y_{f,k} = Cx_{f,k} + Du_{k} + v_{k} \tag{3.60}
\]

For the backward filter, the model is rearranged to progress backwards in time:

\[
x_{b,k-1} = A^{-1}x_{b,k} + A^{-1}Bu_{k-1} + A^{-1}w_{k-1} \tag{3.61}
\]

\[
y_{b,k} = Cx_{b,k} + Du_{k} + v_{k} \tag{3.62}
\]

The Kalman filter is then applied to both forward and backward models. Once
forward and backward estimates at each time step are obtained, they are combined
into a single smoothed estimate, \( \hat{x}_{s,k} \), according to:

\[
\hat{x}_{s,k} = M_{k}\hat{x}_{f,k}^+ + N_{k}\hat{x}_{b,k}^- \tag{3.63}
\]

where matrices \( M \) and \( N \) can be designed so that the smoothed estimate minimizes
the mean squared estimate error. An unbiased estimate requires that \( N_{k} = I_d - M_{k} \).

After substituting, the smoothed estimate is:

\[
\hat{x}_{s,k} = M_{k}\hat{x}_{f,k}^+ + (I_d - M_{k})\hat{x}_{b,k}^- \tag{3.64}
\]
So the goal is to find the matrix $M$ so as to minimize the smoothed estimate error covariance $P_{s,k}$, i.e. when:

$$
\frac{\partial \text{Tr}(P_{s,k})}{\partial M_k} = 0
$$

(3.65)

Solving this for $M$ using the definition of covariance, gives:

$$
M_k = P_{b,k}(P_{f,k} + P_{b,k})^{-1}
$$

(3.66)

and the smoothed estimate covariance is:

$$
P_{s,k} = (P_{f,k}^{-1} + P_{b,k}^{-1})^{-1}
$$

(3.67)

Again, as an example the Kalman smoother is applied to the single particle model in Fig. 3.5 to estimate normalized concentration, using model generated pseudo-data. Fig. 3.6 shows the trace of the estimate covariance matrices for the forward, backward, and smoothed estimates. Clearly, the smoothed estimate results in a lower overall estimate covariance than for either the forward or backward estimates.

Since it is required that the forward and backward estimates be independent, none of the information from the forward filter can be used in the backward filter. Therefore, the initial backward estimate covariance, $P_{b,N}^{-1}$, must be infinite. There are several adjustments that can be made to the backward Kalman filter equations to avoid this initialization problem as well as improve the computational efficiency, such as using an information filter [88], an adjoint filter [14,26], or the Rauch, Tung, and Striebel filter [26,88]. The adjoint filter (also known as the Bierman or modified Bryson-Frazier filter) will be outlined here, as it is computationally efficient and can be easily extended to nonlinear systems. Adjoint variables $\lambda$ and $\Lambda$ are used in the backward filter, which are analogous to the state $\hat{x}_b$ and covariance $P_b$, respectively. The adjoint variables are predicted and corrected starting from the final time step.
Figure 3.5: Example showing forward, backward, and smoothed estimates of normalized concentration using the single particle model to generate pseudo-data.

$k = N$, progressing backward in time. The algorithm equations are outlined in Table 3.3.

Another advantage of the Kalman smoother is that it can be more robust to model error than the standard EKF. For example, consider the same single particle example, where a 50% error in the electrode area is introduced in the model assumed in the smoother. The smoothed estimate shown in Fig. 3.7, while not perfect, tends to be more accurate than either the forward or backward estimates. The effects of the model error get somewhat averaged out in the smoothed estimate.
Table 3.3: Kalman smoother algorithm for state estimation of a generic nonlinear system with assumed process and sensor noise.

<table>
<thead>
<tr>
<th>Filter</th>
<th>Step</th>
<th>Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic</td>
<td></td>
<td>( x_k = f(x_{k-1}, u_{k-1}) + w_{k-1} )</td>
</tr>
<tr>
<td>Nonlinear</td>
<td></td>
<td>( y_k = g(x_k, u_k) + v_k )</td>
</tr>
</tbody>
</table>

### Forward Initialization
\[
\begin{align*}
\hat{x}^+_f,0 &= \mathbb{E}(x_{f,0}) \\
P^+_f,0 &= \mathbb{E}[(x_0 - \hat{x}^+_f,0)(x_0 - \hat{x}^+_f,0)^T] \\
\hat{x}^+_f,k &= f(\hat{x}^+_{f,k-1}, u_{k-1}) \\
y^+_f,k &= g(\hat{x}^+_{f,k}, u_k) \\
P^-_{f,k} &= A_k P^+_f,k-1 A_k^T + Q \\
A_k &= \frac{\partial f}{\partial x} \bigg|_{\hat{x}^+_f,k} \\
K_k &= P^-_{f,k} C_k (C_k P^-_{f,k} C_k^T + R)^{-1} \\
C_k &= \frac{\partial g}{\partial x} \bigg|_{\hat{x}^+_f,k} \\
P^+_{f,k} &= (I_d - K_k C_k) P^-_{f,k} \\
\end{align*}
\]

### Prediction Step

### Correction Step

### Backward Initialization
\[
\begin{align*}
\Lambda^-_N &= -C_N^T[C_N P^-_{f,N} C_N^T + R]^{-1}[y_N - g(\hat{x}^-_{f,N}, u_N)] \\
\Lambda^-_N &= C_N^T[C_N P^-_{f,N} C_N^T + R]^{-1} C_N \\
\lambda^+_k &= A_k^T \lambda^-_{k+1} \\
\Lambda^+_k &= (A_k - I_d)^T \Lambda^-_{k+1} + \Lambda^-_{k+1} A_k \\
\lambda^-_k &= [I_d - C_k^T K_k^T] \lambda^+_k \\
- & -C_k^T[C_k P^-_{f,k} C_k^T + R]^{-1}[y_k - g(\hat{x}^-_{f,k}, u_k)] \\
\Lambda^-_k &= [I_d - K_k C_k^T] \Lambda^+_k [I_d - K_k C_k] \\
+ & +C_k^T[C_k P^-_{f,k} C_k^T + R]^{-1} C_k \\
\end{align*}
\]

### Correction Step

### Smoothing
\[
\begin{align*}
\hat{x}_{s,k} &= \hat{x}^-_{f,k} - P^-_{f,k} \Lambda^-_k \\
P_{s,k} &= P^-_{f,k} - P^-_{f,k} \Lambda^-_k P^-_{f,k} \\
\end{align*}
\]
3.5 Combined State and Parameter Estimation

For certain applications it is desirable to estimate an unknown model parameter, in addition to the model states. This may be a parameter that is unmeasurable (even offline) or a parameter that is expected to change over time. For example in the single particle model, the parameter for the active material volume fraction, $\epsilon_i$, is expected to decrease with age. Estimating time varying model parameters as they change generally helps to keep the model more accurate than if the parameters were treated as constants or scheduled based on some assumed (open-loop) trajectory. A
Figure 3.7: Forward, backward, and smoothed normalized concentration estimates when a 50% error in the electrode area is introduced.

A straightforward way to implement combined state-parameter estimation is to treat the parameters to be estimated as additional states in the model. The state vector is then appended to include the new model states, and the state space model is updated accordingly. Often, if the parameters are not expected to change rapidly, no dynamics other than process noise are assumed for the states representing the parameters. In other words, the parameter states are only updated in the correction step using feedback from the output measurements. For example, take the simple state space model:

$$\dot{x}_1 = \alpha x_1 + u$$  \hspace{1cm} (3.68)
where $\alpha$ is some parameter. If it is now desired to estimate $\alpha$ along with the state $x_1$, the state vector is expanded to: 

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ \alpha \end{bmatrix}.$$ 

The dynamics now become:

$$\dot{x}_1 = x_2 x_1 + u$$

$$\dot{x}_2 = 0$$

This approach is called joint estimation [73]. Notice how this system that was linear becomes nonlinear once the parameter is added as a state. In fact this is almost always the case, as making a state out of any parameter that shows up in the A, B, C, or D matrices of a linear system will either result in a state multiplied by another state or a state multiplied by an input. To avoid the system becoming nonlinear, it is possible to implement two separate observers, one for the states and one for the parameters. This is called dual estimation [73]. The update and correction steps for the states and parameters are handled separately at each time step, with some exchange of information between the two observers. For the simple example in Eq. (3.68), implementing dual state and parameter observers would result in two one-state linear systems. This has advantages in that both systems remain linear and, generally speaking, the observer equations for two one-state models can be computed faster than for one two-state model. The computational difference is small with such a low number of states, but it can make a difference with much larger state models. For the single particle model used in this work, the base model is already nonlinear and only a few parameters are ever estimated so differences in the joint and dual estimation approaches are minor.

Special care must be taken to correctly tune the observers in a joint or dual estimation algorithm. If the parameters are expected to change much more slowly than the states (as is the case for SOH related parameters), the process noise covariance,
\( Q \), for the parameter states should be tuned appropriately to avoid large corrections made to the parameter estimates.

Finally it should be made clear that observability of the original system does not necessarily translate into observability of the new appended system. A separate observability analysis should be conducted to see if inclusion of a parameter in the state vector causes the system to become unobservable.

### 3.6 Summary

Model-based estimation is a common problem applicable to many dynamic systems, where it is desired to estimate the value of certain model states that may not be directly measurable. Whether or not it is possible to find a unique state estimate based on the available measurements depends on the observability properties of the system model. Observability of the model states implies that they can be uniquely reconstructed based on the measurements, and well-established tests exist for determining observability of linear systems. For nonlinear systems, the question of proving uniqueness over all possible values of states and outputs is less clear, and often a local approach is required. This involves calculation of Lie derivatives of the output and testing for local weak observability, that guarantees a unique state solution within some local region. Testing the observability of the single particle model reveals that the states are indeed locally observable under certain conditions (i.e. the cell is not at steady state). That said, the poor degree of observability using the condition number of the observability matrix indicates that it may be difficult to design an observer for all the states of the complete battery model.

There are many methods for state estimation of nonlinear systems, including the extended Kalman filter, particle filter, and unscented Kalman filter. In addition, smoothing techniques provide an opportunity to obtain improved state estimates at
the expense of obtaining them in real time. Still, the focus for online SOC/SOH estimation should be on observers that are computationally cheap.
Chapter 4

State of Charge Estimation

In this chapter, the state estimation techniques described in Chapter 3 are applied to estimate battery state of charge in the reduced-order electrochemical model.

4.1 Introduction

A battery’s state of charge (SOC) indicates how much useful energy is remaining in a given cell or pack. Clearly, this is an important metric to track in an electrified vehicle application. The SOC is used to give a miles-to-empty indicator to the driver in the case of a PHEV or BEV. Additionally, an accurate SOC is a necessary input to the vehicle’s battery management system in order to calculate the optimal power split, correctly balance cells, and facilitate aggressive use of the battery while maintaining safe operation [20, 71, 71, 78]. Unfortunately there is no practical sensor that can be used to measure SOC in the same way that the remaining fuel in a gas tank is easily measured. Instead, SOC must be inferred or estimated from other measurable quantities like current and voltage.
4.1.1 Defining state of charge

First, the concept of state of charge should be properly defined. In general terms, SOC refers to the amount of useful stored charge remaining in a cell, as a percentage of its maximum possible stored charge [79]. However, SOC can be defined differently depending on the context. For battery experiments, the cell SOC is usually calculated using the method of Coulomb (or charge) counting, given by:

\[
SOC_{\text{cell}} = SOC_{\text{cell},0} - \frac{1}{3600C} \int_0^{t_f} I(t)dt
\]  

(4.1)

where \(SOC_{\text{cell},0}\) is the initial SOC at the start of the test, \(C\) (Ah) is the cell capacity between two voltage limits (as defined in Eq. (1.4)), and \(t_f\) (s) is the time at the end of the test. The initial \(SOC_{\text{cell},0}\) is usually obtained by using a lookup table or functional fit between SOC and the cell OCV. When the cell is at rest, the measured cell voltage is equal to the cell OCV, and the corresponding SOC can be determined through the lookup table. SOC is a dimensionless quantity that varies from 0 to 1 between the voltage limits, since the current integration over the test (which gives Ah) is normalized by the total cell capacity (also in Ah).

In the context of electrochemical models, SOC takes on a more physical meaning. Fundamentally, the concept of SOC is related to the transport of lithium from one electrode to the other. A cell with a highly lithiated negative electrode and relatively delithiated positive electrode should correspond to a high SOC, since this implies that the cell has most of its stored charge left to give. In the single particle model (or the porous electrode model in general), an SOC can be defined for each electrode as:

\[
SOC_i = \frac{c_i}{c_{\text{max},i}}
\]  

(4.2)
where $c_i$ is the concentration of lithium intercalated in the spherical active material particle of electrode $i$ and $c_{\text{max},i}$ is the maximum possible concentration of intercalated lithium (a material property). The maximum concentration is related to the specific capacity of the active material by:

\[
c_{\text{max},i} = \frac{3600 \rho_i q_i}{n_e F}
\]  

(4.3)

where $\rho_i$ (kg/m$^3$) is the active material density and $q_i$ (Ah/kg) is the material specific capacity.

Of course the meaning of SOC in this context depends on the choice of $c_i$, since the lithium concentration varies spatially in the electrode. In the reduced-order SPM, the logical choices are to use the bulk particle concentration, $\bar{c}_i$, or the surface concentration, $c_{s,i}$, since these are the only two concentrations predicted by the model. In this way, a bulk and surface SOC can be defined for each electrode as:

\[
\bar{\text{SOC}}_i = \frac{\bar{c}_i}{c_{\text{max},i}}
\]  

(4.4)

\[
\text{SOC}_{s,i} = \frac{c_{s,i}}{c_{\text{max},i}}
\]  

(4.5)

Both the bulk and surface SOC are important metrics to track. The bulk SOC is analogous to the Coulomb counting method, except it counts the total lithium entering or exiting the electrode particles rather than the flow of electrons. Therefore, an estimate of bulk SOC can be validated with experimentally obtained SOC from Coulomb counting. Note that the bulk SOC defined by Eq. (4.4) is not yet directly comparable to the SOC from Coulomb counting in its current form. This is due to the differences in how $C$ and $c_{\text{max}}$ are defined. The procedure for comparing the two metrics is outlined in section 4.2.3.

The surface SOC has implications in the available power than can be extracted at a
given time [20], since the redox reactions are limited by the reactant concentrations at the particle surface. The bulk and surface SOC converge to the same value at steady state, but it can be useful to track both metrics during dynamic current profiles.

4.1.2 Methods for state of charge estimation

The Coulomb counting method is straightforward to implement in the lab, but it suffers in vehicle applications for three practical reasons. First, there may be a sensor bias in the current measurement. Integrated over time, this error accumulates leading to an ever increasing error in the calculated SOC. Often sensor errors are higher onboard a vehicle than in a well controlled laboratory with high-end equipment. Also in a lab setting, there are opportunities to reset the “counter” (and therefore the error accumulation) by forcing the cell to rest and come to equilibrium at periodic points in the testing. Second, there may be an error in the calculation of the initial $SOC_{cell,0}$. This could come from a voltage sensor error, resulting in the incorrect OCV-SOC lookup value, or if the cell has not yet reached equilibrium and the measured cell voltage is not equal to the OCV. Onboard a vehicle, the resting times in between charging or discharging cannot be controlled, and an equilibrium condition may not be reached for some time. Third, the cell capacity may not be known accurately, as it changes over time. In the lab, it is possible to perform periodic checks of the capacity by completely charging and discharging the cell; this is clearly not possible in a vehicle. For these reasons, the Coulomb counting method alone is not a practical way to monitor SOC onboard a vehicle, thereby motivating the need for closed-loop SOC estimation.

Model-based SOC estimation is typically done using either an equivalent circuit model or a reduced-order electrochemical model, the former being the approach used in the automotive industry today. In a circuit model, the Coulomb counting equation
is used to model SOC, which is then used to schedule the circuit parameters. The SOC in the Coulomb counting equation can be treated as a state that is estimated through feedback from the measured cell voltage [24,38,41,71–73]. To reduce the uncertainty regarding the cell capacity, the capacity parameter \( C \) can also be estimated through a joint or dual estimation algorithm [55,71–73,80].

SOC estimation using an electrochemical model amounts to estimating the states associated with the bulk and/or surface concentrations of one or both electrodes [20,29,33,51,64,76,82–84,90]. The advantages and challenges of using electrochemical models for SOC estimation over traditional circuit models are outlined in Section 2.8. In short, electrochemical models have the potential to provide more accurate and informative SOC estimates, that remain more accurate over the cell life.

In this chapter, the single particle model outlined in Chapter 2 is used to estimate SOC of a composite LMO-NMC cell. The calculation of the current split in the composite electrode and the conversion of the composite electrode concentration estimates to a cell level SOC that can be validated experimentally are outlined. The model observability as it relates to observer design is discussed, followed by a discussion of methods to improve observer performance by making additional assumptions in the model. Finally, the SOC estimation algorithm is applied to experimentally collected data.

4.2 SOC Estimation for a Composite Electrode

Due to their popularity in automotive applications, the SOC estimation in this work will focus on cells with a composite electrode, specifically the composite LMO-NMC electrode (though the methods could be used for any combination of electrode active materials). Furthermore, examples of online SOC estimation for composite electrode cells using electrochemical models are absent in the literature, so this is a novel
contribution.

4.2.1 Experimental methods

The experimental methods used to validate the SOC estimation algorithm and the algorithm to estimate loss of cyclable lithium in Chapter 5 are outlined in this section.

Aging campaign

Current and voltage data were taken from cells that were part of an aging campaign under the Center for Automotive Research Industrial Consortium at Ohio State. The tested cells were 15 Ah (nominal) pouch cells with a composite LMO-NMC cathode and graphite anode, manufactured by LG Chem Ltd. The nominal 1 s resistance was 1.75 mΩ. The pouch cells were fixtured in between two steel plates, with Peltier junctions attached to each plate in order to maintain a constant temperature of 30°C during cycling. The cells were cycled using the charge depleting (CD) current profile defined by the United States Advanced Battery Consortium (USABC) [1], which is representative of a PHEV application. The profile, shown in Fig. 4.1, consists of a series of constant current pulses and rests, which are scaled so as to gradually deplete the cell (averaged over a cycle). The magnitude of pulses was determined using a battery size factor (BSF) of 94, representative of a cell in a moderately sized PHEV battery pack. This results in a maximum current pulse of almost 10C (142 A). The CD profile is repeated multiple times (typically 5-7 times) until a lower SOC cut-off is reached. The cells are then charged using a constant-current, constant-voltage (CCCV) charging protocol, where the cutoff current during the constant voltage period is set to C/3. Both the lower SOC cutoff and the constant-current charging rate were varied for different cells, as part of the design of experiments for the aging campaign. Periodic capacity and resistance assessments were conducted at approximately
every 2000-5000 Ah of throughput. The capacity assessments involved an initial 1C CCCV charge to 4.15 V, followed by a 1C discharge to 2.8 V, followed by a second 1C CCCV charge. The measured capacity was taken as the average of the discharge and charge capacities. Current and voltage data were collected with an Amrel SPS 8-150 power supply (paired with an Amrel PLA800-60-300 load).

The results of the aging campaign and a description of the design of experiments is given in Section 5.3.

Model parameterization

Parameterizing the model involved a combination of direct measurements (such as the cell dimensions), values taken from literature, and tuning parameters to fit the experimental data. The parameterization process relies on work done in [60] and [59] for LiFePO$_4$ and LMO-NMC cells, respectively. The anode open circuit voltage was
obtained experimentally by fabricating a coin cell from a disassembled full cell, and
cycling at C/20. The model structure requires separate parameters for each active
material in the composite cathode, so using a fabricated half cell was not an option for
the cathode. Instead, the OCVs of pure LMO and NMC were taken from literature.
Parameters like electrode/separador thicknesses and area could be measured directly
from the deconstructed cell. Kinetic and diffusion related parameters were taken
from literature. Internal resistance was tuned to fit the measured ohmic voltage drop
corresponding to a change in current. The volume fractions of each active material
were tuned to fit constant current cycling of half cells of the anode and cathode. A
low rate was used (C/10) to do the fitting so that the effect of the resistance related
parameters was minimized. Similarly, the electrode composition ranges (defined in
Section 4.2.3) were established at the beginning of life by fitting to the measured full
cell voltage during low, constant current cycling. The model parameters and their
sources are shown in Tables 4.1-4.2.

Note that certain parameters like rate constants, diffusion coefficients, and con-
ductivity may be functions of temperature, in which case they can be scheduled based
on an Arrhenius relationship [60]. However, the temperature variation of parameters
was not a primary concern in this work, since the experiments used to validate the
estimation algorithms were done at a constant 30°C.

The cell voltage model fit is shown in Figs. 4.2-4.3 for a constant 1C discharge
and for the CD profile described previously.

4.2.2 Calculating the current split

The single particle model for a cell with a composite electrode is outlined in Chapter
2. The graphite negative electrode is treated as a single particle, but the positive
electrode consists of two particles (LMO and NMC), acting in parallel. The voltage
Figure 4.2: Measured voltage and predicted voltage during a 1C constant current discharge.

Figure 4.3: Measured voltage and predicted voltage for the charge depleting current profile used in the aging campaign.
Table 4.1: Model parameters used for the composite LMO/NMC electrode.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>Electrode cross-sectional area</td>
<td>0.877 m$^2$</td>
<td>Measured</td>
</tr>
<tr>
<td>$R_{\text{cell}}$</td>
<td>Cell level contact resistance</td>
<td>$1.113 \times 10^{-3}$ Ω</td>
<td>Tuned</td>
</tr>
<tr>
<td>$L_p$</td>
<td>Cathode thickness</td>
<td>$75 \times 10^{-6}$ m</td>
<td>Measured</td>
</tr>
<tr>
<td>$U_{\text{LMO}}$</td>
<td>LMO open circuit voltage</td>
<td>Fig. 4.7</td>
<td>**</td>
</tr>
<tr>
<td>$R_{\text{LMO}}$</td>
<td>LMO particle radius</td>
<td>$2 \times 10^{-6}$ m</td>
<td>*</td>
</tr>
<tr>
<td>$D_{\text{LMO}}$</td>
<td>LMO diffusion coefficient</td>
<td>$6 \times 10^{-15}$ m$^2$/s</td>
<td>*</td>
</tr>
<tr>
<td>$k_{\text{LMO}}$</td>
<td>LMO rate constant</td>
<td>$1.9 \times 10^{-9}$ m$^{2.5}$/mol$^{0.5}$/s</td>
<td>*</td>
</tr>
<tr>
<td>$R_{l,\text{LMO}}$</td>
<td>LMO contact resistance</td>
<td>$1.4 \times 10^{-4}$ Ω</td>
<td>Tuned</td>
</tr>
<tr>
<td>$\epsilon_{\text{LMO}}$</td>
<td>LMO volume fraction</td>
<td>0.32</td>
<td>*</td>
</tr>
<tr>
<td>$\eta_{\text{LMO}}$</td>
<td>LMO specific capacity</td>
<td>105 Ah/kg</td>
<td>* + Tuned</td>
</tr>
<tr>
<td>$\rho_{\text{LMO}}$</td>
<td>LMO density</td>
<td>4220 kg/m$^3$</td>
<td>*</td>
</tr>
<tr>
<td>$y_{0,\text{LMO}}$</td>
<td>LMO SOC corresponding to the upper voltage limit</td>
<td>0.211</td>
<td>Tuned</td>
</tr>
<tr>
<td>$y_{f,\text{LMO}}$</td>
<td>LMO SOC corresponding to the lower voltage limit</td>
<td>0.990</td>
<td>Tuned</td>
</tr>
<tr>
<td>$U_{\text{NMC}}$</td>
<td>NMC open circuit voltage</td>
<td>Fig. 4.7</td>
<td>#</td>
</tr>
<tr>
<td>$R_{\text{NMC}}$</td>
<td>NMC particle radius</td>
<td>$2 \times 10^{-6}$ m</td>
<td>*</td>
</tr>
<tr>
<td>$D_{\text{NMC}}$</td>
<td>NMC diffusion coefficient</td>
<td>$6.5 \times 10^{-15}$ m$^2$/s</td>
<td>*</td>
</tr>
<tr>
<td>$k_{\text{NMC}}$</td>
<td>NMC rate constant</td>
<td>$2.5 \times 10^{-10}$ m$^{2.5}$/mol$^{0.5}$/s</td>
<td>*</td>
</tr>
<tr>
<td>$R_{l,\text{NMC}}$</td>
<td>NMC contact resistance</td>
<td>$1.4 \times 10^{-4}$ Ω</td>
<td>Tuned</td>
</tr>
<tr>
<td>$\epsilon_{\text{NMC}}$</td>
<td>NMC volume fraction</td>
<td>0.18</td>
<td>*</td>
</tr>
<tr>
<td>$\eta_{\text{NMC}}$</td>
<td>NMC specific capacity</td>
<td>158 Ah/kg</td>
<td>* + Tuned</td>
</tr>
<tr>
<td>$\rho_{\text{NMC}}$</td>
<td>NMC density</td>
<td>4770 kg/m$^3$</td>
<td>*</td>
</tr>
<tr>
<td>$y_{0,\text{NMC}}$</td>
<td>NMC SOC corresponding to the upper voltage limit</td>
<td>0.115</td>
<td>Tuned</td>
</tr>
<tr>
<td>$y_{f,\text{NMC}}$</td>
<td>NMC SOC corresponding to the lower voltage limit</td>
<td>0.978</td>
<td>Tuned</td>
</tr>
</tbody>
</table>

* [59]
** [50]
# [87]
Table 4.2: Model parameters used for the negative electrode and electrolyte.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>(U_n)</td>
<td>Anode open circuit voltage</td>
<td>Fig. 5.15</td>
<td>*</td>
</tr>
<tr>
<td>(L_n)</td>
<td>Anode thickness</td>
<td>(50 \times 10^{-6}) m</td>
<td>Measured</td>
</tr>
<tr>
<td>(R_n)</td>
<td>Anode particle radius</td>
<td>(3.5 \times 10^{-6}) m</td>
<td>*</td>
</tr>
<tr>
<td>(D_n)</td>
<td>Anode diffusion coefficient</td>
<td>(1.5 \times 10^{-14}) m(^2)/s</td>
<td>*</td>
</tr>
<tr>
<td>(k_n)</td>
<td>Anode rate constant</td>
<td>(1 \times 10^{-8}) m(^{2.5})/mol(^{0.5})/s</td>
<td>* + Tuned</td>
</tr>
<tr>
<td>(\epsilon_n)</td>
<td>Anode active material</td>
<td>volume fraction</td>
<td>0.52</td>
</tr>
<tr>
<td>(q_n)</td>
<td>Anode specific capacity</td>
<td>372 Ah/kg</td>
<td>*</td>
</tr>
<tr>
<td>(\rho_n)</td>
<td>Anode density</td>
<td>2250 kg/m(^3)</td>
<td>*</td>
</tr>
<tr>
<td>(x_0)</td>
<td>Anode SOC corresponding</td>
<td>to the upper voltage limit</td>
<td>0.880</td>
</tr>
<tr>
<td>(x_f)</td>
<td>Anode SOC corresponding</td>
<td>to the lower voltage limit</td>
<td>0.097</td>
</tr>
<tr>
<td>(L_{sep})</td>
<td>Separator thickness</td>
<td>(20 \times 10^{-6}) m</td>
<td>Measured</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>Liquid activity coefficient</td>
<td>1.75</td>
<td>#</td>
</tr>
<tr>
<td>(\epsilon_e)</td>
<td>Separator porosity</td>
<td>0.52</td>
<td>**</td>
</tr>
<tr>
<td>(D_e)</td>
<td>Liquid diffusion coefficient</td>
<td>(3.78 \times 10^{-11}) m(^2)/s</td>
<td>**</td>
</tr>
<tr>
<td>(c_{e,nom})</td>
<td>Nominal liquid concentration</td>
<td>1000 mol/m(^3)</td>
<td>**</td>
</tr>
<tr>
<td>(t^+)</td>
<td>Transference number of Li</td>
<td>0.39</td>
<td>**</td>
</tr>
<tr>
<td>(\sigma_e)</td>
<td>Conductivity of liquid</td>
<td>12.34 S/m</td>
<td>#</td>
</tr>
</tbody>
</table>

* [59]  
** [60]  
# [93]  

across the two parallel paths is the same, but the current is split between each path according to the current split factor, \(\beta_i\). For convenience, the electrode state space dynamics are repeated here in discrete time for the positive electrode:

\[
x_{i,k} = A_{d,i}x_{i,k-1} + B_{d,i,k-1}u_{k-1}
\] (4.6)
where $i$ indicates either LMO or NMC. For the positive electrode, $\beta_{i,k}$ must be solved at each time step to determine how the current is allocated, with the constraint that the currents going to each particle sum to the total current (i.e. $\beta_{LMO,k} + \beta_{NMC,k} = 1$).

The $\beta_{i,k}$ term appears nonlinearly in the output equation, so one possible solution method is to guess the value of each $\beta_{i,k}$, calculate each active material potential (Eq. (4.7)), and iterate until the two material potentials converge to the same value. This approach could be implemented using a standard nonlinear solver, such as the Newton Raphson method [98]; however, this may not be ideal for online estimation since it requires iteration at each time step. Alternatively, an approximated solution is obtained more quickly using the following linearization procedure:

1. At each time step, $k$, advance the states and concentrations forward in time using the finite difference method:

$$x_{i,k} = A_{d,i}x_{i,k-1} + B_{d,i,k-1}u_{k-1}$$

$$\begin{bmatrix} c_{s,i,k} \\ \bar{c}_{i,k} \end{bmatrix} = \begin{bmatrix} a_{0,i} & a_{1,i} & a_{2,i} \\ a_{0,i} & a_{0,i}b_{2,i} & a_{0,i}b_{3,i} \end{bmatrix} x_{i,k}$$

$$\phi_{i,k} = U_i(c_{s,i,k}) - \eta_i(c_{s,i,k}, \beta_{i,k}u_k) - R_{c,i}\beta_{i,k}u_k$$  \hspace{1cm} (4.7)

where the current splits in $B_{d,i,k-1}$ are from the previous time step.

2. Construct a set of three (non)linear equations, two for the output potentials of
each particle and one enforcing that the current splits sum to 1.

\[
f_1 = \phi_{p,k} = U_{LMO}(c_{s,LMO,k}) - \eta_{LMO}(c_{s,LMO,k}, \beta_{LMO,k} I_k) - R_{c,LMO} \beta_{LMO,k} I_k
\]

\[
f_2 = \phi_{p,k} = U_{NMC}(c_{s,NMC,k}) - \eta_{NMC}(c_{s,NMC,k}, \beta_{NMC,k} I_k) - R_{c,NMC} \beta_{NMC,k} I_k
\]

\[
f_3 = I_k = \beta_{LMO,k} I_k + \beta_{NMC,k} I_k
\]

where \( \phi_p \) is the total potential of either particle.

3. Remove the nonlinear dependence on \( \beta_i \) in \( f_1 \) and \( f_2 \) by taking a Taylor series expansion about the current splits at the previous time step:

\[
\phi_{p,k} \approx f_1 \bigg|_{\beta_{LMO,k-1}} + \frac{\partial f_1}{\partial \beta_{LMO}} \bigg|_{\beta_{LMO,k-1}} (\beta_{LMO,k} - \beta_{LMO,k-1})
\]

\[
\phi_{p,k} \approx f_2 \bigg|_{\beta_{NMC,k-1}} + \frac{\partial f_2}{\partial \beta_{NMC}} \bigg|_{\beta_{NMC,k-1}} (\beta_{NMC,k} - \beta_{NMC,k-1})
\]

4. Finally, the current splits and potential can be written as the solution to a
matrix inversion problem:

\[
\begin{bmatrix}
\frac{\partial f_1}{\partial \beta_{LMO}} |_{\beta_{LMO,k-1}} & 0 & -1 \\
0 & \frac{\partial f_2}{\partial \beta_{NMC}} |_{\beta_{NMC,k-1}} & -1 \\
I_k & I_k & 0 \\
\end{bmatrix}
\begin{bmatrix}
\beta_{LMO,k} \\
\beta_{NMC,k} \\
\phi_{p,k} \\
\end{bmatrix}
= 
\begin{bmatrix}
-f_1|_{\beta_{LMO,k-1}} + \frac{\partial f_1}{\partial \beta_{LMO}} |_{\beta_{LMO,k-1}} \beta_{LMO,k-1} \\
-f_2|_{\beta_{NMC,k-1}} + \frac{\partial f_2}{\partial \beta_{NMC}} |_{\beta_{NMC,k-1}} \beta_{NMC,k-1} \\
I_k \\
\end{bmatrix}
\]

(4.11)

In addition to the current splits, solving (4.11) also gives a value for the linearized output electrode potential, \(\phi_{p,k}\); however, this value is discarded at this point. Instead the original nonlinear output equation (4.7) is used to calculate the electrode potential, thereby avoiding the voltage prediction inaccuracies seen with the linearization process. In this way, the errors associated with the output linearization only affect the calculation of \(\beta_i\), and not the calculation of the output voltage (apart from the indirect effect from \(\beta_i\)).

A comparison of the calculated current split between the iterative Newton-Raphson method and the linearization method is shown in Fig. 4.4. The dynamic, charge depleting current profile followed by a CCCV charge described in Section 4.1 is used to make the comparison. The two methods compare very well and are virtually indistinguishable. The improvement in computational effort using the linearization method depends on the stopping criteria for the iterative Newton-Raphson method. However, at a given time step, the computational effort for one iteration of the Newton-Raphson method is comparable to the linearization method (both require a 3x3 matrix inversion), so the linearization method will always be as fast or faster than the Newton-Raphson.
Figure 4.4: Comparison of the calculated current split for each composite material, between the iterative Newton-Raphson method and the non-iterative linearization method. Subplot b.) shows a close-up of the initial pulses.
Figure 4.5: Comparison of the calculated current going to each composite material particle, between the iterative Newton-Raphson method and the non-iterative linearization method. Subplot b.) shows a close-up of the initial pulses.
One possible drawback of the electrode model with parallel particles is that for instances when the current abruptly changes in magnitude and/or direction, the calculated $\beta_i$ for a given particle can greatly exceed 1 (resulting in a large negative $\beta_i$ for the other particle). While it is unclear if this occurrence makes physical sense, it typically lasts for only a few time steps before the $\beta_i$s return to more sensible values between 0 and 1. Furthermore, these are usually instances when the total current is near zero, so a large $\beta_i$ does not translate into a large current going to a given particle. Figure 4.5 shows the magnitude of the current going to each particle, plotted along with the total applied current. The instances when the current split $\beta_i$ exceeds 1 do not result in a large current applied to one particle. In fact, there are no instances when the current going to a given particle exceeds the total current by more than a few Amps (in these cases, this lasts only a few time steps).

It should also be noted that if the state estimate has not yet converged, there will be some error in the calculation of $\beta_i$, since it relies on the current state estimate. This would in turn cause an error in the state estimate calculated at the next time step. However, this does not affect the stability of the state estimate. Even if there is an extreme error in $\beta_i$, eventually the concentration estimate of one of the materials will become saturated, which will decrease $\beta_i$ for that material, thereby driving the estimate back towards the correct value. Instead, calculation of $\beta_i$ using an unconverged state estimate can be categorized as a model error, which may affect the estimate accuracy, but will not cause the estimate to become unstable. Furthermore, in a real automotive application, there are periodic instances when the current is near zero, and the influence of $\beta_i$ is negligible. During these times the state estimate can converge accurately, thereby minimizing the error in $\beta_i$ once the current resumes.

To summarize the solution method for the composite positive electrode: at each time step, $k$, the states are updated in discrete time using the calculated values for $\beta_{i,k-1}$ at the previous time step (Eq. 4.6), the values for $\beta_{i,k}$ at the current time step
are then calculated using the linearization method outlined above (Eq. (4.11)), and finally the output voltages are calculated using the updated states and pre-calculated values for $\beta_{i,k}$ (Eq. (4.7)).

### 4.2.3 Calculating a cell level SOC

In order to compare with the Coulomb counting method, the bulk SOC estimate of a given electrode must be converted into a comparable cell level SOC. The term cell level SOC is used here referring to the fact that this SOC is a property of the complete cell operating between two voltage limits. This requires careful consideration of how each SOC metric is defined. First, the SOC from Coulomb counting, or cell level SOC, ranges from 0 to 1 corresponding to the lower and upper voltage limits, respectively. The bulk electrode SOC, $\bar{SOC}$ defined in Eq. (4.4), ranges from 0 to 1 corresponding to a completely delithiated electrode and a fully saturated electrode, respectively. These two SOC ranges are not the same, since the voltage limits do not correspond to a fully lithiated or delithiated electrode. For safety reasons, the cell voltage limits correspond to an “almost” lithiated or delithiated electrode. The bulk SOCs that correspond to the upper and lower voltage limits are defined as:

\[
x_0 : \text{Negative electrode bulk SOC corresponding to the upper voltage limit.}
\]
\[
x_f : \text{Negative electrode bulk SOC corresponding to the lower voltage limit.}
\]
\[
y_0 : \text{Positive electrode bulk SOC corresponding to the upper voltage limit.}
\]
\[
y_f : \text{Positive electrode bulk SOC corresponding to the lower voltage limit.}
\]

These SOC ranges can be visualized in Fig. 4.6.

After defining these parameters, the bulk concentration of either electrode can be
Figure 4.6: Definition of the electrode composition ranges corresponding to the upper and lower cell voltage limit.
converted to an overall cell SOC by:

\[ SOC_{cell} = 1 - \frac{S\bar{O}C_p - y_0}{y_f - y_0} = \frac{S\bar{O}C_n - x_f}{x_0 - x_f} \]  (4.13)

where \( S\bar{O}C_n = \frac{\bar{c}_n}{c_{max,n}} \). For the composite positive electrode, the \( S\bar{O}C_p \) must be defined slightly differently, since the concentrations in each composite active material are generally different. This is done by summing the moles of lithium in each composite active material to compute the total moles of lithium in the composite electrode, \( N_{Li,p} \), and the total maximum moles of lithium, \( N_{Li,max,p} \):

\[ N_{Li,p} = \bar{c}_{LMO} AL_p \epsilon_{LMO} + \bar{c}_{NMC} AL_p \epsilon_{NMC} \]  (4.14)

\[ N_{Li,max,p} = c_{max,LMO} AL_p \epsilon_{LMO} + c_{max,NMC} AL_p \epsilon_{NMC} \]  (4.15)

Then, the overall positive electrode SOC is simply the moles of lithium divided by the maximum moles of lithium:

\[ S\bar{O}C_p = \frac{N_{Li,p}}{N_{Li,max,p}} = \frac{\bar{c}_{LMO} \epsilon_{LMO} + \bar{c}_{NMC} \epsilon_{NMC}}{c_{max,LMO} \epsilon_{LMO} + c_{max,NMC} \epsilon_{NMC}} \]  (4.16)

The calculation of \( S\bar{O}C_n \) could be done in terms of moles of lithium as well, but for the single material negative electrode, the ratio of concentrations is the same as the ratio of lithium (since the active material volume in the numerator and denominator cancel out). Following a similar process, the overall positive SOC corresponding to the upper and lower voltage limits can be defined in terms of the bulk SOC of each active material:

\[ y_0 = \frac{y_{0,LMO} c_{max,LMO} \epsilon_{LMO} + y_{0,NMC} c_{max,NMC} \epsilon_{NMC}}{c_{max,LMO} \epsilon_{LMO} + c_{max,NMC} \epsilon_{NMC}} \]  (4.17)
Figure 4.7: LMO and NMC OCVs as a function of normalized concentration. Initial and final SOC are defined at the upper and lower cell voltage limits. The overall initial and final SOC, $y_0$ and $y_f$ correspond to a cell SOC (top axis) of 1 and 0, respectively.

$$y_f = \frac{y_{f,LMO}c_{max,LMO} + y_{f,NMC}c_{max,NMC}}{c_{max,LMO} + c_{max,NMC}}$$ (4.18)

Finally, once $\bar{SOC}_p$, $y_0$, and $y_f$ are defined, the cell SOC can be calculated via Eq. (4.13). The conversion of the positive electrode SOC estimates to a cell level SOC can be visualized in Fig. 4.7. The OCVs of each material give the initial and final $y_{0,i}$ and $y_{f,i}$, defined at the voltage limits. These are then converted to an overall $y_0$ and $y_f$ via Eqs. (4.17)-(4.18).

Quantifying the bulk electrode composition ranges at the voltage limits, $x_0$, $x_f$, $y_0$, and $y_f$, is not entirely straightforward, absent a reference electrode in the cell. Essentially there is not a unique set of $x_0$, $x_f$, $y_0$, and $y_f$ that correspond to the upper and lower cell voltage limits. However, if these parameters are defined at an
equilibrium condition, then there is a unique set of $y_0$ and $y_f$ for a given $x_0$ and $x_f$, and vice versa. This is due to the fact that at equilibrium, the cell voltage limit equals to the difference between positive and negative electrode OCVs. So for example, a given $x_0$ corresponds to a unique $U_n$ at equilibrium. If the upper cell voltage limit is $V_{\text{high}}$, then the positive OCV is uniquely defined as $U_p = V_{\text{high}} + U_n$. Finally, $y_{0_{\text{LMO}}}$, $y_{0_{\text{NMC}}}$, and $y_0$ are all uniquely defined from $U_p$. In this way, a given $x_0$ corresponds to a unique $y_0$, and vice versa. The same is true for $x_f$ and $y_f$ at the lower voltage limit. The reason this works is that the electrode SOC-OCV curves are all one-to-one and onto. Still, neither parameter pair $x_0$ and $x_f$ nor $y_0$ and $y_f$ are known values. Therefore, they must be determined by fitting the model to experimental data. This is done using a low rate, constant current discharge (C/10) between the voltage limits. A least-squares algorithm is then used to vary, say $x_0$ and $x_f$ until the error between the predicted voltage and measured voltage is minimized. At the low rate, the sources of model error for kinetic and diffusion processes are minimal, and the accuracy of the resulting parameter estimates depend mostly on the accuracy of each electrode material OCV.

To summarize, the bulk SOC of each electrode corresponding to the voltage limits are first determined by fitting the model to experimental data. As the bulk electrode concentrations are estimated by an observer, they are converted to a cell level SOC via Eq. (4.13). At this point, the cell SOC estimate can be directly compared with the SOC from Coulomb counting.

### 4.2.4 Initializing the electrodes

The model states must be initialized at the start of a given current profile. Although a main feature of an observer is that it is robust to initial condition errors, the initial guess should be close to the true value if possible. This is particularly important
when implementing the “open-loop” and “steady-state” methods described in the next section. Assuming the cell begins a given profile at equilibrium, the initial cell level SOC calculated from each electrode should be equal, according to Eq. (4.13). In other words:

\[
\bar{\text{SOC}}_{p,0} - y_f \frac{y_0 - y_f}{x_0 - x_f} = \bar{\text{SOC}}_{n,0} - x_f
\]

(4.19)

where estimates of the electrode composition ranges, \(x_0, x_f, y_0,\) and \(y_f,\) are assumed to be available. Also, the initial measured cell voltage must equal the difference between the positive and negative OCV, evaluated at their respective SOC:

\[
V_0 = U_{i,0}(\bar{\text{SOC}}_{i,0}) - U_{n,0}(\bar{\text{SOC}}_{n,0})
\]

(4.20)

In the case of the composite positive electrode, both the LMO or NMC OCV may be used (forming two separate equations). Finally, a nonlinear iterative solver is used to solve for the initial SOC of each electrode active material (3 equations, 3 unknowns).

### 4.3 Observability of the Complete Model and Possible Improvements

In Chapter 3, the single particle model that includes the states of both electrodes is shown to be locally observable with respect to the cell voltage measurement. The resulting observability matrix has full rank, however it is very poorly conditioned. The condition number of a matrix is a measure of how close the matrix is to losing rank, and can be used as a measure of “degree of observability” [28,52]. It is calculated by:

\[
\kappa(A) = ||A^{-1}|| \cdot ||A||
\]

(4.21)
A high condition number indicates a poorly conditioned matrix and model states that are close to becoming unobservable. If this is the case, an observer may have difficulty obtaining an accurate state estimate in the presence of noise and initial condition errors [28,52].

Computing the matrix condition number analytically is not practical for the single particle model due to the complexity of the matrix elements, however it is possible to evaluate the matrix numerically by plugging in typical values for parameters, states, and inputs over a given drive cycle. The result is that the observability matrix when the states of both electrodes are included has a very high condition number ($\sim 10^{10}$) over the CD-CCCV profile, as shown in Fig. 4.14. There is not a hard rule that the observability matrix must be below some threshold for the model to be suitable for observer design. A perfectly conditioned matrix has a condition number of 1, but this is rarely achievable. Therefore, it is more meaningful to analyze how this condition number may change by making various assumptions in the model, as is done in the following sections.

It is possible for the observability matrix to be poorly conditioned simply because of large differences in the scaling of the states (for example if one state was in km and another in mm). Therefore, to obtain a meaningful metric for the degree of observability, the condition number should technically be normalized to remove the effect of unit scaling [75]. The normalizing term can be defined as the condition number of the observability matrix when every state is assumed to be measured (resulting in a “C” matrix equal to the identity matrix). This essentially represents the ideal case from the perspective of observability. If there are any ill-effects from poor scaling, they will be canceled out through this normalization. That said, the value of this normalization factor in the case of the single particle model is relatively small (around 1.8), which means the high condition number is not due to scaling issues but is intrinsic to the model structure.
Figure 4.8: Surface SOC estimate results when attempting to estimate the states of both electrodes using an EKF. A perfect model with added sensor noise is assumed in order to generate pseudo-data, and an initial condition error on the positive states is imposed.

Figure 4.8 shows the result when attempting to implement an EKF to estimate all 9 of the electrode states (3 for each active material). For this demonstration, the single particle model is used to generate current and voltage “pseudo-data”. Noise is then added to the pseudo-data, which is then used as the measurement for feedback in the EKF. This is essentially the best case scenario, as it does not consider any model error. The CD profile outlined in Fig. 4.1 is repeated 7 times to simulate a PHEV or BEV application. An initial condition error is imposed on the positive electrode states. Clearly, the state estimates are not able to recover from this initial condition error in a reasonable time, and the surface SOC estimates of each active material are very inaccurate. Note that the result is similar if an initial condition error is imposed on the negative electrode states (let alone on both electrodes). This

\[^1\text{These estimate results are for one particular tuning of the noise covariances to try and get a fast convergence of the estimates within the allotted time. Alternate tuning does show the ability to converge to the true value over the course of many such cycles, as shown in Chapter 5.}\]
Figure 4.9: Block diagram of the “open-loop” algorithm for estimating the positive electrode states. The negative electrode and liquid phase are simulated open-loop in order to generate a pseudo-measurement of the positive electrode potential. Alternatively, the negative electrode states could be estimated by simulating the positive electrode open-loop.

motivates one of two approaches to strengthen the observability of the model, by making further assumptions.

4.3.1 Approach 1: Open-loop estimation of one electrode

The first approach is to simulate one of the two electrodes open-loop. In other words, eliminate the “correction step” for the states of one electrode, in which the states are corrected using feedback from the output measurement. As an example, consider simulating the negative electrode states open-loop, and applying an observer to the positive states only. The predicted negative electrode states can be used to obtain an open-loop prediction of the negative electrode voltage, which can then be subtracted from the measured cell voltage (along with the liquid voltage) to obtain a pseudo-measurement of the positive electrode voltage. This is of course not a real measurement, but one that relies on the open-loop predictions. The positive electrode pseudo-voltage can then be used as feedback in the observer to estimate the positive electrode states. This process is outlined in Fig. 4.9.

The open-loop method is demonstrated in Fig. 4.10 and 4.11, where the same CD
Figure 4.10: Surface SOC estimate results when using the open-loop method, where the states of only the positive electrode are estimated using an EKF. A perfect model with added sensor noise is assumed in order to generate pseudo-data, and an initial condition error on the positive states is imposed.

A profile is used to generate voltage and current pseudo-data. In Fig. 4.10, the EKF is applied only to the positive electrode states (simulating the negative electrode states open-loop), and an initial condition error is imposed on the positive states. Conversely in Fig. 4.11, the EKF is applied only to the negative electrode states (simulating the positive electrode states open-loop), and an initial condition error is imposed on the negative states.

Using this open-loop method, the state estimates converge from the imposed initial condition error and remain accurate for the remainder of the simulation. This is true if the state feedback is applied to the positive or negative electrode states. However, a major assumption here is that the electrode that is simulated open-loop is initialized correctly. If this is not the case, then the estimate accuracy will suffer, as an open-loop estimate cannot overcome an initial condition error (unlike a closed-loop estimate).
Figure 4.11: Surface SOC estimate results when using the open-loop method, where the states of only the negative electrode are estimated using an EKF. A perfect model with added sensor noise is assumed in order to generate pseudo-data, and an initial condition error on the negative states is imposed.

4.3.2 Approach 2: Steady-state approximation for one electrode

A second approach to improve the model observability is to develop a relationship between the bulk concentration of one electrode and the bulk concentration of the other electrode, using conservation of lithium [29]. Neglecting lithium consuming side reactions, conservation of lithium means that for every lithium atom and electron leaving an electrode during a charge or discharge, a lithium atom and electron should enter the opposite electrode. For a single drive cycle, neglecting the effect of side reactions is a good assumption. The relationship between each electrode bulk SOC
This represents an additional constraint on the estimated model states. This constraint is already enforced in the model itself, as the bulk concentrations at each electrode are tied to the same current input, but there is no guarantee that it remains enforced once the states are “corrected” by the observer. It is possible that the states of one electrode will be corrected more (in terms of lithium) than the states of the other electrode, effectively violating conservation of lithium.

There are a number of ways to enforce conservation of lithium while still allowing for state correction from the measurement. One method is to design the observer gains so that conservation of lithium is maintained [51], although this no longer allows for the gains to be determined by, say the EKF algorithm. A variation on this is to apply an observer to correct the states of one electrode, and then apply an equal and opposite correction to the states of the other electrode, so that lithium is conserved. Finally, the approach taken in [29] is to neglect the concentration dynamics in one electrode and use Eq. (4.22) to relate the electrode bulk concentrations. For example
consider the procedure in Fig. 4.12, where the negative electrode concentration dynamics are neglected, as they are generally faster than the positive electrode. The 3rd order Padé model of the positive electrode can be used normally to compute the positive surface and bulk concentrations. Then the negative electrode bulk concentration is calculated from the positive bulk concentration using Eq. (4.22). However, since the concentration gradient in the negative particle is neglected, the negative electrode output voltage must be evaluated at the bulk concentration, rather than the surface concentration. This introduces a small error in the overall cell voltage prediction, and in particular, may have an adverse effect on estimating available power.

These various methods are all similar in that they essentially remove a degree of freedom in the estimation of the states of one electrode by enforcing that the states of this electrode must be linked to the states of the opposite electrode at all times. In this work, the approach outlined in Fig. 4.12 is used, since it is compatible with any type of observer and it is more straightforward to understand from an observability perspective.

This method, referred to as the “steady-state” method, is implemented in Fig. 4.13 for the CD profile, under the assumption of a perfect model. An initial condition is imposed on the positive electrode states, and the concentration estimates are able to converge to the true values. Note that for the negative electrode, only the bulk SOC is estimated, so the estimate does not track the true surface SOC dynamics; however, the estimate converges to the true value during periodic rests.

An improvement to this method could be to use an algebraic relationship like the polynomial approximation to obtain the negative electrode surface concentration from the bulk concentration, thereby maintaining the negative electrode concentration overpotential; although the improvement is minimal due to small value of this overpotential and the relatively inaccurate frequency response of the polynomial approximation.
Figure 4.13: Surface SOC estimate results when using the steady-state method, where the positive electrode states are estimated using an EKF, and the negative electrode bulk concentration is obtained via conservation of lithium. A perfect model with added sensor noise is assumed in order to generate pseudo-data, and an initial condition error on the positive states is imposed.

4.3.3 Observability of open-loop and steady-state methods

Clearly the open-loop and steady-state methods result in much better observer performance than using the full model to estimate the states of both electrodes. This fact can be predicted by analyzing the observability matrix condition number for each method. The state dynamics and output for each method are shown in Table 4.3, and the observability matrix for each method can be obtained using the same process of computing Lie derivatives outlined in Chapter 3. An observability matrix need only be computed considering the states of either active material in the composite electrode (not both together), since both parallel particles have the same output voltage.
Table 4.3: Model equations for estimating the states of both electrodes, estimating the states of one electrode using the open-loop method, or estimating the positive states and using the steady-state approximation to obtain the negative concentration. The OCVs in the output can be replaced by a cubic spline section, as in Section 3.2.3, to facilitate calculation of the Lie derivatives.

| Estimating both electrode states (i = LMO, NMC) |  
| $\dot{x} = f(x, u) = Ax + B(t)u$ |  
| $A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1/b_{3,i} & -b_{2,i} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1/b_{3,i} & -b_{2,i} \end{bmatrix}$ | $B(t) = \begin{bmatrix} 0 \\ 0 \\ (\beta_i(t)R_n) \frac{1}{b_{3,i}} \\ 0 \\ 0 \\ -\left( R_n \frac{1}{3FAL_{p\epsilon_i}} \right) \frac{1}{b_{3,n}} \end{bmatrix}$ |  
| $\left[ \begin{array}{c} c_{s,i} \\ c_{i} \\ c_{s,n} \\ c_{n} \end{array} \right] = \left[ \begin{array}{cccc} a_{0,i} & a_{1,i} & a_{2,i} & 0 \\ a_{0,i} & a_{0,i}b_{2,i} & a_{0,i}b_{3,i} & 0 \\ 0 & 0 & 0 & a_{0,n} \\ 0 & 0 & 0 & a_{0,n} \end{array} \right] x$ |  
| $V = g(x, u) = U_i(c_{s,i}) - \eta_i(c_{s,i}, \beta_i(t)u) - R_{c,i}\beta_i(t)u - \bar{U}_n(c_{s,n}) - \eta_n(c_{s,n}, u) - R_{c,n}u$ |  

| Open-loop method (i=LMO, NMC, n) |  
| $\dot{x}_i = f(x_i, u) = A_i x_i + B_i(t)u$ |  
| $A_i = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ | $B_i(t) = \begin{bmatrix} 0 \\ 0 \\ (\beta_i(t)R_n) \frac{1}{b_{3,i}} \end{bmatrix}$ |  
| $\left[ \begin{array}{c} c_{s,i} \\ c_{i} \end{array} \right] = \left[ \begin{array}{cccc} a_{0,i} & a_{1,i} & a_{2,i} \\ a_{0,i} & a_{0,i}b_{2,i} & a_{0,i}b_{3,i} \end{array} \right] x_i$ |  
| $\phi_i = g(x_i, u) = U_i(c_{s,i}) \pm \eta_i(c_{s,i}, \beta_i(t)u) \pm R_{c,i}\beta_i(t)u$ |  

| Steady-state method (i=LMO, NMC) |  
| $\dot{x}_i = f(x_i, u) = A_i x_i + B_i(t)u$ |  
| $A_i = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ | $B_i(t) = \begin{bmatrix} 0 \\ 0 \\ (\beta_i(t)R_n) \frac{1}{b_{3,i}} \end{bmatrix}$ |  
| $\left[ \begin{array}{c} c_{s,i} \\ c_{i} \end{array} \right] = \left[ \begin{array}{cccc} a_{0,i} & a_{1,i} & a_{2,i} \\ a_{0,i} & a_{0,i}b_{2,i} & a_{0,i}b_{3,i} \end{array} \right] x_i$ |  
| $\tilde{c}_n = f(\tilde{c}_i)$ |  
| $V = g(x_i, u) = U_i(c_{s,i}) - \eta_i(c_{s,i}, \beta_i(t)u) - R_{c,i}\beta_i(t)u - \bar{U}_n(\tilde{c}_n) - \eta_n(\tilde{c}_n, u) - R_{c,n}u$ |  

157
The resulting observability matrix condition numbers are shown in Fig. 4.14, over the course of the CD profile used previously, with the addition of a CCCV charge at the end. For this plot, the composite electrode materials are treated separately; in other words, the positive electrode model parameters from either LMO or NMC are used.

It is evident that overall both the open-loop method and the steady-state method decrease the condition number by about 4 orders of magnitude. The improvement in condition number is also about the same between the open-loop and steady-state methods. Note that the condition number for the case of estimating both electrode states is undefined (equal to infinity) during the initial 150 s rest period, when the current is zero and the cell is at equilibrium. This was mentioned in Section 3.2.3, as the observability matrix loses rank under these conditions. However, this is not the case for the open-loop or steady-state methods, as the resulting observability matrix does not lose rank even at equilibrium.

To understand why the open-loop and steady-state methods greatly lower the observability condition number and improve observer performance, it is helpful to understand what characteristics of the system cause the observability matrix of the full cell model to be poorly conditioned in the first place. To demonstrate this, consider a linearized version of the 3rd order Padé model that includes the states of both electrodes (assume a single material cathode). In this linear model, the OCV can be considered a linear function of concentration and the charge transfer overpotential can be considered Ohmic. It is shown in Chapter 3 that these linear approximations render the model unobservable. Determining the reason for this can give some insight into why the original nonlinear model is poorly observable. The “A” and “C” matrices
Figure 4.14: Log of the observability matrix condition number, plotted over the course of a CD-CCCV profile. In a.) the states of both electrodes are assumed to be estimated, in b.) the states of only one electrode are estimated (the other is simulated open-loop), and in c.) the positive electrode states are estimated, with the negative electrode bulk concentration obtained via the steady-state method.
of the linear model have the following structure:

\[
A = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & a_1 & a_2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & a_3 & a_4 \\
\end{bmatrix}
\]

\[
C = \begin{bmatrix}
c_1 \\
c_2 \\
c_3 \\
c_4 \\
c_5 \\
c_6 \\
\end{bmatrix}
\]

where \(a_i\) and \(c_i\) are simply place holders denoting a unique matrix element. An alternate test for observability of a linear system (called the Popov-Belevitch-Hautus (PBH) test) is to compute the matrix:

\[
\begin{bmatrix}
\lambda I_d - A \\
C
\end{bmatrix}
\]

(4.24)

The system is observable if and only if this matrix is full rank for any \(\lambda\). Note that the only values of \(\lambda\) that can cause a loss of rank are eigenvalues of matrix \(A\) (by the Cayley-Hamilton Theorem). Analyzing this matrix allows insight into which eigenvalues (or modes) of matrix \(A\) might contribute to the loss of rank. The PBH
It is easy to see that this matrix loses rank for the eigenvalue \( \lambda = 0 \) (allows columns 1 and 4 to be written as linear combinations of each other). In fact there are two eigenvalues at zero, one associated with each electrode. Another way to see this is in the transfer function relating surface concentration for a given electrode to input current (Eq. (2.97)), which contains a pure integrator \( \frac{1}{s} \). The two eigenvalues at zero (or two poles at \( s = 0 \)) associated with the two electrodes are therefore the cause of the model unobservability. It is not entirely surprising that a linear system consisting of two integrators, with a measurement that contains the sum or difference of the states, is unobservable. For example, consider a tank of water being filled by two separate hoses, but only the total volume of water in the tank can be measured. The volume of water added by each hose represents the two integrators. In this case, there would be no way to uniquely determine the volume of water being added by each individual hose. However, take one hose away, and now the problem becomes trivial. This is essentially what is being done by implementing the open-loop or steady-state methods in the single particle model. The open-loop and steady-state methods effectively remove the states of one electrode from the model that is applied in the observer. The open-loop method simply removes the states of one electrode with the
assumption that they are known and need not be estimated, thereby removing one of the zero eigenvalues from the A matrix. The steady-state method lets the entire cell model be defined by the positive states alone (since the negative concentration is obtained algebraically from the positive states), again removing one of the zero eigenvalues. Applying either method to the linear single particle model in Eq. (4.23), turns the originally unobservable system into an observable system. Therefore, it is not surprising that application of the open-loop or steady-state methods improves the observability of the nonlinear single particle model.

Of course the improvement in observability comes at a cost. In the open-loop method, the electrode that is simulated open-loop cannot correct for initial condition errors. Likewise in the steady-state method, the relationship between bulk concentrations of either electrode relies on accurate knowledge of the composition ranges $x_0$, $x_f$, $y_0$, and $y_f$, which may be uncertain. Without a reference electrode in the cell, these may simply be unavoidable concessions in order to design a practical SOC estimator.

### 4.4 Validation of SOC Estimates

The open-loop and steady-state methods described previously are now implemented in an EKF using experimental data taken from a 15 Ah LMO-NMC pouch cell. A CD-CCCV current profile is imposed on the cell, with a charge rate of 3C/2. Results are shown in Fig. 4.15-4.16 using the steady-state method, where an initial condition error is imposed on the positive electrode states. The cell SOC estimate converges quickly from an imposed 25% initial condition error, and shows good agreement with the SOC from Coulomb counting, with a maximum error of less than 2% during the CD portion of the cycle. Increased estimate error up to 4% is seen during the constant voltage charge due to poor open-loop model agreement during this portion
of the charge. This level of SOC estimate error is similar among all of the cells tested. Fig. 4.16 shows the estimated surface and bulk concentrations of each individual composite electrode material over the same PHEV cycle, as well as the current split factor; although again, these cannot be easily validated experimentally. However, the general trend of LMO material taking on more of the initial current until it becomes saturated is consistent with an X-ray diffraction study of the LMO-NMC composite electrode [66]. The ability of this model to estimate the surface concentration of each active material is particularly advantageous in estimating available power. For example, once the LMO active material becomes saturated near the end of the CD portion of the profile, the discharge power of the cell will be significantly lowered. At this point, the high power LMO material can no longer take on any more lithium and the overall cell power will be limited by the kinetic and diffusion properties of the lower power NMC material.

The cell SOC estimate error is shown in Fig. 4.17, comparing the open-loop and steady-state methods. The two methods obtain very similar results. However, the two methods do differ if there are imposed errors in the electrode composition ranges, $x_0$, $x_f$, $y_0$, and $y_f$. Errors in these parameters will negatively affect both methods. The composition ranges affect the initialization of the open-loop electrode states in the open-loop method, and the calculation of the negative electrode bulk concentration in the steady-state method. Fig. 4.18 shows the method comparison when a 15% error is imposed on the $x_0$ parameter. Note that the error in $x_0$ translates to an error in $y_0$ as well, since the two are related. Although the open-loop method generally has a lower SOC estimate error during the first stages of the CD cycle, the error gets much larger near the end of the CD cycle and during the charge. It appears that overall, the SOC estimation is slightly more robust to errors in the composition range using the steady-state method.
Figure 4.15: Estimation results over a CD, CCCV charge profile. a) Imposed current profile, b) predicted and measured cell voltage, c) cell SOC estimate and SOC from Coulomb counting, showing convergence from an initial condition error, and d) cell SOC estimate error.
Figure 4.16: Estimation results for cell 2 over a CD, CCCV charge profile. a) Estimated bulk and surface normalized concentrations for each composite material, b) current split factor, $\beta_{LMO}$ ($\beta_{NMC}$ is simply equal to $1 - \beta_{LMO}$), and c) the current (in A) going to the LMO particle compared to the total current.
Figure 4.17: Cell SOC estimate error, comparing open-loop and steady-state methods.

4.5 Summary and Future Work

Battery SOC is an essential input to a vehicle’s battery management system in order to calculate miles-to-empty, determine the available power to/from the battery, calculate the optimal power split in a hybrid, balance cells within a pack, and ensure that the battery operates safely. Traditional SOC estimation involves the use of an equivalent circuit model in which the Coulomb counting equation is employed together with feedback from the voltage measurement. However, SOC estimation using electrochemical models offers several advantages over circuit models due to their
Figure 4.18: Cell SOC estimate error, comparing open-loop and steady-state methods, with a 15% error imposed on the $x_0$ parameter.

ability to preserve the physicality of the model parameters. However, the poor observability of the complete battery model introduces challenges when attempting to estimate SOC. Therefore, approximations can be made, including making open-loop estimates of the states of one electrode or applying the principle of lithium conservation in order to improve the model observability. This improvement in observability comes at the expense of introducing potential initial condition errors that cannot be corrected with feedback.

Specifically for the composite electrode cells, the current split between each com-
posite active material must be calculated at each time step. A linearization procedure is outlined to accomplish this with minimal computational effort. The composite electrode also requires converting the concentration estimates of each active material obtained by the single particle model to a cell level SOC that can be validated experimentally. Applying the SOC estimation algorithm using an extended Kalman filter to experimentally generated voltage data over a typical drive cycle results in good agreement between estimated and measured SOC.

Future work could include further validation of the SOC estimation algorithm. In particular, experimental validation of the individual active material SOC estimates in the composite electrode would be a valuable contribution. This could be done with X-ray diffraction or other techniques for material characterization. Validating the individual SOCs of the LMO and NMC materials would also act as an indirect validation of the current split calculation, as well as a validation of the parallel particle model structure in general. Finally, prediction of available power using the composite electrode model structure could be investigated. A comparison between the available power prediction using the parallel particles to represent the composite electrode versus a lumped single particle approach could provide additional motivation for using the parallel particle model structure.
Chapter 5

State of Health Estimation

Up to this point, the state of charge estimation problem was addressed assuming constant model parameters. In reality, some of these parameters may change as the battery degrades over time. This chapter focuses on designing and validating algorithms to track these model parameters throughout the life of the battery.

5.1 Introduction

Battery degradation, or aging, is an unfortunate side effect inherent to any electrochemical storage device. While the development of advanced electrode materials aims to reduce this degradation, the life of today’s batteries is typically shorter than the life of most other vehicle components, motivating the need to track the degradation process over the battery life and understand how it affects performance. To this end, battery state of health (SOH) estimation is an important topic of study in the design of battery management systems for electrified vehicles. From the perspective of a vehicle BMS, there are two clear goals for any SOH estimation scheme: to provide an accurate miles-to-empty or SOC metric to both the driver (particularly in the case of a BEV) and the BMS, and to provide the maximum discharge and charge power...
that can be requested by the BMS. An additional goal may be to give an indication to the driver if a cell/module/pack is badly degraded and needs replacement, or to predict when this may occur.

5.1.1 Defining state of health

In general terms, state of health refers to the current performance of a battery, normalized by the performance when the battery was new. This definition is vague because there is no agreed upon industry standard for how to define SOH. The performance metric that is used to define SOH may be different, depending on the application. In fact, any battery parameter that changes with usage can be used as an indicator for SOH [79]. Common metrics of interest for a BMS are capacity and internal resistance, in which case SOH could be defined as: \( SOH = \frac{C_{\text{current}}}{C_{\text{new}}} \) or \( SOH = \frac{R_{\text{new}}}{R_{\text{current}}} \). However, even these terms are somewhat ambiguous, as capacity and internal resistance depend on a given testing procedure used to quantify them. For example, capacity may be a function of C-rate, temperature, and current direction, while internal resistance may be a function of SOC, temperature, and current direction.

In the context of model-based estimation, SOH can take on a more well-defined meaning, referring to the current value of a model parameter normalized by the parameter value at BOL; although even this can lead to multiple definitions of SOH, depending on the parameter of interest. Therefore, apart from introducing the concept, the term SOH is generally avoided in this chapter; instead, the SOH estimation problem is framed as a model parameter estimation problem, where a change in a given model parameter corresponds to a specific degradation mechanism. Specifically, estimation of the parameters associated with loss of electrode active material and loss of cyclable lithium is investigated in this chapter. Estimating these parameters, in conjunction with the battery model itself, provides a framework for obtaining the
miles-to-empty or SOC metric and maximum power that are the ultimate goals of the BMS.

5.1.2 Methods for state of health estimation

The state of health estimation problem can be approached in a variety of ways. Data-driven (or black box) approaches do not require the use of a battery model, instead relying on a large collection of data that can be used to map the change in battery performance metrics to measurable environmental factors [79]. These include neural network or fuzzy logic techniques; however, these approaches typically require a large number of experimental data to provide a meaningful answer, since the lack of a model limits their ability to adapt to conditions that haven’t already been experienced. Model-based approaches rely on some form of battery model, in which some of the model parameters are expected to change as the battery ages. These model parameters can then be tracked over time using any number of estimation techniques. Traditionally, model-based SOH estimates are made with the use of equivalent circuit models, in which parameters like capacity and resistance are estimated. However, as stated in Section 2.8, retention of physically meaningful parameters can give electrochemical models several advantages over circuit models.

There have been a few attempts to use electrochemical based models for online SOH estimation using dual SOC/parameter estimation. In [76], the active material fractions of each electrode are estimated along with the bulk and surface SOC, using an unscented Kalman filter (UKF), though the results were not validated experimentally in the presence of model error. SOH estimation was done in [63, 65, 85, 86] by applying least squares techniques to estimate model parameters associated with aging; however, the computational effort required by these approaches may limit their applicability to onboard estimation. In general, electrochemical model-based SOH
estimation algorithms that are validated with experimental aging data are scarce in the literature, and to the author’s knowledge, no such studies have been conducted on cells with composite electrodes. Composite electrodes add to the complexity of the electrochemical model, particularly if it is desired to estimate the SOC and SOH of each electrode material individually.

This chapter focuses on the estimation of parameters associated with two prominent degradation mechanisms, loss of electrode active material and loss of cyclable lithium, and the experimental validation of these estimates. Each degradation mechanism is first approached individually, in which only one mechanism is assumed to occur, followed by an investigation considering both mechanisms occurring simultaneously.

5.2 Estimating Active Material Loss

In this section, estimation of capacity loss through loss of active electrode material is investigated. Loss of active material either through a dissolution process or through cracking of the conductive binder can be a significant source of capacity in many Li-ion chemistries [5, 47, 50, 56, 95]. Loss of active material is a particular concern at later stages of life or at high temperatures [32]. Although it is possible to lose active material of both electrodes, the focus in this work is on loss of active material within the composite positive electrode. The composite LMO-NMC electrode is particularly vulnerable to dissolution of manganese [11, 31, 32]. In addition, the composite electrode is the more challenging electrode to design an observer for, so validating the estimation algorithm for the composite electrode is a more rigorous validation than for the single material anode. In the context of the single particle model, estimating loss of active material translates to estimating the model parameters $\epsilon_{LMO}$ and $\epsilon_{NMC}$, which represent the total active material volume fractions of each material. These two
parameters are treated as additional states in the model, in a joint state-parameter estimation algorithm. Therefore, the complete state vector consists of 3 states that combine to give the surface and bulk concentration of the LMO particle \((c_{s,LMO} \text{ and } \bar{c}_{LMO})\), 3 states that combine to give the surface and bulk concentration of the NMC particle \((c_{s,NMC} \text{ and } \bar{c}_{NMC})\), and 2 states representing the active material volume fractions of LMO and NMC \((\epsilon_{LMO} \text{ and } \epsilon_{NMC})\). The modified positive electrode state dynamics become:

\[
\dot{x}(t) = \begin{cases}
    x_2(t) \\
    x_3(t) \\
    -\frac{1}{b_3,LMO}x_2(t) - \frac{b_2,LMO}{b_3,LMO}x_3(t) + \frac{R_{LMO}}{3b_3,LMO FAL_p x_7(t)}\beta_{LMO}(t)u(t) \\
    x_5(t) \\
    x_6(t) \\
    -\frac{1}{b_3,NMC}x_5(t) - \frac{b_2,NMC}{b_3,NMC}x_6(t) + \frac{R_{NMC}}{3b_3,NMC FAL_p x_8(t)}\beta_{NMC}(t)u(t) \\
    0 \\
    0
\end{cases}
\]  

(5.1)

\[
\Phi_i(t) = U_i(c_{s,i}(t)) - \eta(c_{s,i}(t), \epsilon_i(t), \beta_i(t)u(t)) - R_{c,i}\beta_i(t)u(t), \quad i = LMO, NMC
\]  

(5.2)

The time derivatives of the two volume fraction states are set to zero, since they are expected to change slowly over time. Note that the once linear electrode state dynamics become nonlinear once the volume fractions are modeled as states (leading to the inverse of a state multiplied by an input).

In order to better understand the mechanism responsible for capacity degradation in the cell, it is useful to decompose the total active material volume fraction
parameters, $\epsilon_{LMO}$ and $\epsilon_{NMC}$, as follows:

\begin{align*}
\epsilon_{LMO} &= \epsilon_p \epsilon_{LMO,\text{split}} \\
\epsilon_{NMC} &= \epsilon_p \epsilon_{NMC,\text{split}} \tag{5.3}
\end{align*}

\[ \epsilon_{LMO,\text{split}} + \epsilon_{NMC,\text{split}} = 1 \]

Here, $\epsilon_p$ represents the overall fraction of active (electrode material) versus inactive (carbon and binder) material. Note that $\epsilon_p$ is independent of whether the electrode material is LMO or NMC. Additionally, $\epsilon_{LMO,\text{split}}$ and $\epsilon_{NMC,\text{split}}$ represent the fractional split of active material between LMO and NMC, defined so that the fractional splits sum to 1. After making this decomposition, it is possible to write $\epsilon_p$, $\epsilon_{LMO,\text{split}}$, and $\epsilon_{NMC,\text{split}}$ as functions of the original volume fraction estimates:

\begin{align*}
\epsilon_p &= \epsilon_{LMO} + \epsilon_{NMC} \\
\epsilon_{LMO,\text{split}} &= \frac{\epsilon_{LMO}}{\epsilon_{LMO} + \epsilon_{NMC}} \tag{5.4} \\
\epsilon_{NMC,\text{split}} &= \frac{\epsilon_{NMC}}{\epsilon_{LMO} + \epsilon_{NMC}}
\end{align*}

The reason for this decomposition is so that, as the original volume fractions are estimated, it is possible to predict the mechanism responsible for the degradation. For example, if a decrease in $\epsilon_p$ is seen without a change in the volume fraction splits, this suggests that the degradation may be a mechanical fracturing of binder that is affecting both LMO and NMC active materials. Alternatively, if a decrease in the total $\epsilon_{LMO}$ is seen without a change in the total $\epsilon_{NMC}$, this suggests a loss of LMO material through dissolution in the electrolyte.
5.2.1 Observer design

It is now desired to estimate the expanded positive electrode state vector, to obtain estimates of both active material concentrations and volume fractions. In this analysis, the EKF is chosen as the observer due to its computational efficiency, using a joint estimation approach (i.e. estimating all of the states together). Alternatively, dual observers could be built to estimate the original states and the volume fraction parameters separately.

Model observability

Expanding the positive electrode state vector to include the active material volume fractions requires an observability analysis to ensure that the states are still observable. The Lie derivative approach outlined in Section 3.2.3 is applied to the new state dynamics (Eq. 5.1) and output (Eq. 5.2). The new volume fraction states, $x_7$ and $x_8$, appear in the scaling terms for the input current and also in the charge transfer overpotential term in the output equation. Note that a pseudo-measurement of the electrode potential is assumed to be available using the open-loop approach outlined in Chapter 4. Once again, the resulting observability matrix can be shown to have full rank, although it is too lengthy to show. However in this case, the observability matrix does lose rank if the current is zero. This fact is easy to see looking at the state equations, as a zero current removes the coupling between the volume fraction states and concentration-related states (in state equations 3 and 6), and eliminates the charge transfer overpotential entirely. This effectively cuts all ties between the volume fraction states and the output, rendering them unobservable.

The condition number of the observability matrix over the course of a CD-CCC profile is shown in Fig. 5.1. The value of the condition number is much higher than for the original one electrode system. In fact, it is of the same order as the condition
number for the case of estimating the concentrations of both electrodes, as shown in Fig. 4.14. One way to interpret the high condition number for the appended system is that the output is relatively insensitive to changes in the volume fractions over much of the SOC range. This implies that it may be difficult to design an observer for estimating volume fractions. However the advantage in this case, compared to the case of estimating both electrode concentrations, is that the volume fractions are expected to change much more slowly than the electrode concentrations. This gives some flexibility in tuning the observer. In particular, the gains applied to the volume fraction states can be tuned down, relative to the gains applied to the concentration-related states, so that they respond much more slowly to an output residual. The tradeoff is that the volume fraction estimates will take much longer to converge from an initial condition error; however, this is an acceptable concession in the context of SOH estimation, where parameters change slowly over many cycles.
Figure 5.2: 1 cycle of the CD-CS-CCCV current profile (a.) used to generate pseudo-data, and the resulting voltage response (b.).
Validating the estimation algorithm in simulation

The problem of estimating active material volume fractions is first tested in simulation. The single particle model is used to generate current and voltage pseudo-data for a charge-depleting profile down to 25% SOC, followed by a charge-sustaining profile and a CCCV charge at 1C. This profile, shown in Fig. 5.2, is representative of a PHEV application. The CD profile is the same as was used in Section 4.4 to cycle the 15 Ah LMO-NMC pouch cells, and the CS profile is simply the CD profile with the pulse magnitudes modified to be charge sustaining ($\int I(t)dt = 0$) over a given cycle. 50 consecutive CD-CS-CCCV cycles are simulated, where the total active material volume fraction of LMO is continuously decreased, as a function of Ah throughput. The results in [32] are used to justify the exponential decay of active material as a function of Ah throughput, as this was observed at later stages of life for these composite LMO-NMC cells. The loss of LMO active material only is intended to be representative of a Mn dissolution process, though certainly NMC could be lost as well. Noise is added to the voltage pseudo-data and the EKF is implemented to estimate the total active material volume fractions. The EKF is implemented using the open-loop approach outlined in Chapter 4, where the negative electrode and liquid phase sub-models are simulated open-loop and the observer is applied only to the positive electrode states. The process noise covariance matrix is tuned as follows, so that the volume fraction states will respond much more slowly than the concentration-related states:

$$Q = 1e^{-18} diag (5, 0.005, 5e^{-6}, 5, 0.005, 5e^{-6}, 5000, 5000) \quad (5.5)$$

Although the number of the noise covariance for the volume fraction states (5000) is actually higher than the covariance for the original states, this is due to the relative
magnitudes of the states. In terms of the estimate response, this particular tuning results in a much slower response for the volume fraction states.

Even with the assumption of a perfect model, the open-loop estimate of the negative electrode will have small initial condition errors if the active material estimates are not accurate. The reason for this is that the active material volume fractions are needed to calculate the electrode composition ranges, $x_0$, $x_f$, $y_0$ and $y_f$, which are then used to initialize the two electrodes. Initial condition errors in the positive electrode can be corrected via output feedback, but the initial condition errors in the negative electrode cannot be corrected and will have some effect on the estimate accuracy. The estimation results are shown in Fig. 5.3, where the decompositions of the total volume fractions into an overall active material volume (independent of LMO/NMC) and a volume fraction split are also shown. The volume fraction estimates are able to converge from an imposed initial condition error, and track the loss of LMO active material over the 2000 Ah of throughput. There is a small bias once the estimates converge from the initial condition error, that can be attributed to initial condition errors in the open-loop negative electrode.

5.2.2 Experimental validation of active material estimates

In order to validate the estimation algorithm experimentally, composite LMO-NMC half cells were built and repeatedly cycled to induce aging. The cell fabrication and testing was done by Kevin Rhodes at Ford Motor Company’s Research and Innovation Center. The fabrication and testing procedures are summarized here.

Cells were built with mass ratios of NMC to LMO of 1:1 and 3:1, and were subject to 20 consecutive C/5 constant current cycles to induce aging. The half cells were built with a pure lithium reference electrode for two main reasons. First, it is desired to have a direct measurement of the composite positive electrode potential in order
Figure 5.3: Volume fraction estimate results showing a.) the total active material volume fractions, b.) the overall active material volume fraction, and c.) the volume fraction splits.
to avoid errors associated with open-loop estimation of one electrode. The pure lithium electrode dynamics are fast (which makes it a good reference electrode), so the cell voltage measurement is a good approximation of just the composite electrode potential. While a direct measurement of electrode potential in this way is not possible in an automotive application, it is desired in this case in order to validate whether the estimates are correctly tracking the degradation.

Second, to best validate the active material estimates, the only source of cell degradation should be loss of active material. The presence of other degradation mechanisms that are not accounted for in the model will produce biased estimates. The pure lithium reference electrode ensures that any loss in capacity could not be due to shortage of cyclable lithium, as it contains excess lithium to overcome any lithium consumed in SEI formation. Other degradation mechanisms could cause a rise in internal resistance that could, in turn, decrease the effective capacity. This has been demonstrated in LMO-NMC composite cells [32], as higher current rates accelerate aging by hindering the cell kinetics, particularly noticeable at later stages of life. Testing at higher current rates may warrant the estimation of resistance-related parameters; however, the increase in internal resistance measured over the 20 cycles was 88% for the 1:1 cell and 13% for the 3:1 cell and, although this resistance rise introduces some model error, it does not contribute significantly to the capacity fade at the C/5 rate. Therefore, almost all of the observed capacity fade can be attributed to loss of active material, and changes in the estimated active material volume fractions can be converted to decreases in capacity and directly compared to the observed capacity fade.

**Experimental methods**

In order to validate the active material volume fraction estimation algorithm, Swagelok-type cells with composite LMO-NMC electrodes were fabricated and cycled to induce
aging.

Free standing composite electrodes composed of 80% active material, 10% conductive additive, and 10% binder were cast using a doctor blade from acetone based slurry. Active materials used were Li(Ni\(_{1/3}\)Mn\(_{1/3}\)Co\(_{1/3}\))O\(_2\) (NMC-111) from BASF and LMO from Alfa Aesar (99.5% metal basis). Vulcan XC72R carbon (Cabot) served as the conductive additive and Kynar HSV900 (Arkema) was used as the binder material. Electrodes were cast with weight ratios of NMC to LMO of 1:1 and 3:1. First, the Kynar was dissolved in acetone (Fischer Scientific, A18-4) for about 15 hours while mixing with a magnetic stir bar. Next, the carbon conductive additive was mixed into the solution which was subsequently sonicated for 10 minutes. The active materials were then mixed into the slurry and the entire mixture was blended for 10 minutes at 3000rpm using as Ross LSR high shear mixer. The slurry was cast onto a clean glass plate and allowed to dry while covered. The resulting free standing electrode tapes were cut into 12 mm diameter disks, weighed on an analytical balance, and dried in a vacuum oven at 100°C for 12 hours prior to being transferred to an argon filled glove box.

Swagelok-type cells were assembled in a glove box using a composite electrode, a 14mm diameter Celgard 2400 separator disk, a 12mm diameter lithium foil disk, and 1.0M LiPF\(_6\) in ethylene carbonate (EC) and dimethyl carbonate (DMC) (3:7 by weight) electrolyte. Once assembled, the cells were removed from the glove box and were put through 20 full charge/discharge cycles controlled by a Solartron Analytical 1470E cell test system. Cycling was conducted using a constant-current constant-voltage (CCCV) protocol where a C/5 rate was used during the galvanostatic phase with voltage limits of 3V and 4.3V. Current was allowed to drop to a C/25 rate before ending each potentiostatic phase.
Figure 5.4: Open loop model fit of a C/5 discharge at BOL for the a.) 1:1 ratio and b.) 3:1 ratio cells.
Model parameterization

Since the experimental validation of the active material estimation algorithm uses half cells with pure lithium negative electrodes that were cycled at a low rate (C/5), the model does not include any dynamics associated with the negative electrode or liquid phase. Essentially, instantaneous kinetics are assumed for the redox reaction at the pure lithium electrode and only ohmic losses are considered in the electrolyte.

For the 1:1 and 3:1 ratio cells, the volume fraction split parameters are quantified at beginning of life (BOL) based on the known weight ratios and densities, resulting in initial values of $\epsilon_{LMO,split} = 0.53$ and $\epsilon_{NMC,split} = 0.47$ for the 1:1 cell and $\epsilon_{LMO,split} = 0.27$ and $\epsilon_{NMC,split} = 0.73$ for the 3:1 cell. The BOL value of the parameter $\epsilon_p$ is quantified by considering the 4:1 ratio of active to inactive (binder and carbon) material and an estimated electrode porosity of 0.4, resulting in an initial $\epsilon_p = 0.8(1 - 0.4) = 0.48$. The OCV curves for LMO and NMC are obtained from C/5 discharges of cells that only contained either LMO or NMC electrode material. In constructing the OCV, the resistive losses at the C/5 rate are accounted for by adding the voltage drop 1 s after current is drawn to the measured voltage. Other model parameters such as diffusion coefficients and rate constants are taken from literature [59], but these parameters in particular do not have a large effect on the model prediction at the low current rates that were tested. The open-loop model prediction of a C/5 discharge at BOL is shown in Fig. 5.4. The model predictions for each cell agree well, with root mean square errors of 17 mV and 9 mV for the 1:1 and 3:1 ratio cells, respectively.

Estimation results and validation

The reduced-order electrochemical model is used in an extended Kalman filter (EKF) in order to estimate the model states, including the two appended states representing the parameters $\epsilon_{LMO}$ and $\epsilon_{NMC}$. Due to the pure lithium anode and negligible liquid
dynamics, the cell voltage measurement acts as a direct measure of the composite electrode potential, resulting in a model whose states are observable from the cell voltage measurement.

In this study, the extended Kalman filter is used as the nonlinear observer, as it computationally cheap and has been used successfully for estimation of battery SOC and SOH [8, 9, 29, 71–73, 76]. Other observer types are investigated in Section 5.3. The EKF consists of a prediction step, where the state vector is advanced in time according to the model dynamics, followed by a correction step, where the states are corrected based on the error between the measured output and the predicted output. Zero-mean, white, Gaussian process and sensor noise are assumed and, as is common practice, the noise covariance matrices are tuned to achieve the desired estimator performance [29, 76].

Figures 5.5-5.6 show the volume fraction estimates for the 1:1 and 3:1 ratio cells over the course of the 20 charge and discharge cycles. The two volume fraction states were only estimated during the discharge portions, due to poor model agreement during charging. In other words, the latest volume fraction estimates are simply held constant for the duration of the charge.

The first point of validation is to determine if the change in the volume fraction estimates over time can predict the decrease in measured cell capacity. The theoretical capacity of the composite electrode, assuming full utilization of each material, can be written as the sum of each material capacity:

$$C = AL_p \epsilon_p \epsilon_{LMO,split} q_{LMO} \rho_{LMO} + AL_p \epsilon_p \epsilon_{NMC,split} q_{NMC} \rho_{NMC} \quad (5.6)$$

where $q_i$ is the material specific capacity (in Ah/kg) and $\rho_i$ is the material density (in kg/m$^3$). This equation could be used to relate changes in each active material volume fraction to changes in capacity. However, since the cell resistance and voltage
Figure 5.5: Active material volume fraction estimation results for the 1:1 ratio cell.
Figure 5.6: Active material volume fraction estimation results for the 3:1 ratio cell.
Figure 5.7: Model-generated lookup table used to convert volume fraction estimates to a cell capacity for the 1:1 ratio cell (3:1 ratio cell not shown). Lines on the figure represent possible combinations of $\epsilon_{\text{LMO}}$ and $\epsilon_{\text{NMC}}$ for a given cell capacity. The two points represent the specific combination of $\epsilon_{\text{LMO}}$ and $\epsilon_{\text{NMC}}$ estimated by the EKF at the beginning and end of the experiment.

limits prevent full utilization of this capacity, a more accurate relationship between volume fractions and actual capacity is established by using the model to simulate changes in $\epsilon_{\text{LMO}}$ and $\epsilon_{\text{NMC}}$, record the resulting capacity change, and construct a lookup table. The result, shown in Fig. 5.7, is a near linear relationship between each active material volume and cell capacity, though the slope differs for LMO and NMC materials. The lines on the figure represent possible combinations of $\epsilon_{\text{LMO}}$ and $\epsilon_{\text{NMC}}$ that give the same cell capacity. Lines are drawn corresponding to the cell capacity at BOL, the capacity after the 20 cycles that were tested, and corresponding to a 20% capacity loss (typically signifying end of life). Points are also shown that give the specific combination of $\epsilon_{\text{LMO}}$ and $\epsilon_{\text{NMC}}$ estimated by the EKF at the beginning and end of the experiment.
Figure 5.8: Estimated cell capacity based on volume fraction estimates for the a.) 1:1 ratio and b.) 3:1 ratio cells, compared to the measured capacity.
This lookup table is then used to convert the volume fraction estimates to capacity estimates. The capacity estimates for the two cells track fairly well with the decreasing measured capacity over the 20 cycles, as shown in Fig. 5.8. The estimate error is less than 3% for the 1:1 ratio cell and less than 5% for the 3:1 ratio cell.

Although the ability of the estimates to predict the capacity fade gives some confidence in the estimation algorithm, validating the exact values of each volume fraction estimate is difficult without a direct measurement of how the volume fractions change as the cell ages. However, it is possible to validate the trends in the volume fraction estimates through the use of a differential capacity analysis. In Figs. 5.5-5.6, the estimation results predict that there is not much change seen in the volume fraction split parameters, $\epsilon_{LMO,\text{split}}$ and $\epsilon_{NMC,\text{split}}$, whereas a clear decrease is seen in the $\epsilon_p$ parameter. In other words, the estimates predict that the degradation mechanism is not affecting one material preferentially over the other. These trends can be confirmed by analyzing changes in differential capacity curves throughout the aging process in order to understand the cause of those changes.

A differential capacity analysis involves constructing a curve of the derivative of voltage with respect to capacity (the curve is inverted in this paper to show $dQ/dV$), and associating changes in the features of the voltage derivative curve to physical changes in the electrode [15–18]. The differential capacity curves for the two cells at various stages of life are plotted in Fig. 5.9. In both cells, there is a clear upward shift in the peaks at 3.7 V, 4 V, and 4.1 V, as the cells age. In order to understand what physical changes could cause this result, the electrochemical model is used to produce differential capacity curves under three different degradation scenarios: loss of LMO active material (decrease in $\epsilon_{LMO}$), loss of NMC active material (decrease in $\epsilon_{NMC}$), and loss of overall active material (decrease in $\epsilon_p$). The model prediction of the differential capacity curves for the 1:1 cell under these degradation scenarios are plotted in Fig. 5.10.
Figure 5.9: Experimental differential capacity curves measured at various stages of life for the a.) 1:1 ratio and b.) 3:1 ratio cells.
Figure 5.10: Model-generated differential capacity curves under different aging scenarios: a.) loss of LMO material, b.) loss of NMC material, and c.) loss of overall active material. Each scenario corresponds to a signature change in the peaks at various voltages.
The loss of LMO case shows an upward shift in the peaks at 4 V and 4.1 V, but not in the peak at 3.7 V. The loss of NMC case shows an upward shift in the peak at 3.7 V, but not in the other peaks. Only the loss of overall active material case shows an upward shift in all three peaks, matching the behavior seen in the experimental curves. This result is consistent with the volume fraction estimates in Figs. 5.5-5.6 that show a decrease in the $\epsilon_p$ parameter, with little change in the $\epsilon_{\text{LMO,split}}$ or $\epsilon_{\text{NMC,split}}$ parameters.

Another way to interpret the differential capacity curves is to look at how the upward shift in the 3.7 V, 4.0 V, and 4.1 V peaks are related to the decrease in overall capacity. In equation 5.6 for the theoretical capacity, the overall volume fraction, $\epsilon_p$, can be factored out of each term, so it follows that capacity scales linearly with $\epsilon_p$. Likewise, the dQ/dV peaks should also scale linearly with $\epsilon_p$, and therefore also with capacity. Figure 5.11 shows the dQ/dV peaks taken from the experimental differential capacity curves plotted against the measured cell capacity. Excluding the first 3 cycles, which may have been affected by the initial cell formation, both cells show a roughly linear relationship between the upward shift in the peaks and decrease in capacity. This linear relationship has a similar slope among the 3.7 V, 4.0 V, and 4.1 V peaks, again indicating that the loss in active material volume fraction is independent of the specific electrode material.

5.3 Estimating Loss of Cyclable Lithium

In this section, estimation of capacity loss due to loss of cyclable lithium is investigated. Unlike with SOC estimation, capacity estimates do not necessarily need to be made in real time, since capacity only changes significantly over months or years of usage. This gives some flexibility in making capacity estimates, and it is possible to wait for ideal conditions before making an estimate. Loss of cyclable lithium can
Figure 5.11: Experimental $dQ/dV$ peaks versus measured cell capacity for a.) 1:1 ratio and b.) 3:1 ratio cells, showing a roughly linear relationship for each peak, following the initial 3 cycles.
occur through side reactions at the solid-electrolyte interface, that consume lithium. This causes a shift in the electrode composition ranges of the positive ($y_0$ to $y_f$) and negative ($x_0$ to $x_f$) electrodes, shown in Fig. 5.12, as the electrodes becomes less lithiated overall. Again, $x_0$ and $y_0$ correspond to the upper voltage limit, while $x_f$ and $y_f$ correspond to the lower voltage limit.

The shift in the electrode composition ranges resulting from a loss in cyclable lithium can also be visualized in Fig. 5.13. The diagram represents the electrode concentrations of each electrode by the shaded areas. Since the cell voltage limits stay consistent through the cell life, a decrease in cyclable lithium during a discharge results in an overall lower $x_f$ and $y_f$. If this cell is now charged to the upper voltage limit, after having lost lithium, the result will be an overall lower $x_0$ and $y_0$.

The normalized concentration operating range of either electrode can be converted to a capacity in Ah by:

$$C = (x_0 - x_f)c_{max,n}FAL_n\epsilon_n/3600$$

$$= \frac{FAL_p}{3600}((y_f,\text{LMO} - y_0,\text{LMO})c_{max,\text{LMO}}\epsilon_{\text{LMO}} + (y_f,\text{NMC} - y_0,\text{NMC})c_{max,\text{NMC}}\epsilon_{\text{NMC}})$$

(5.7)

This is the realizable capacity between the two cell voltage limits and so is identical for each electrode. Therefore, an estimate of the normalized concentration range of a given electrode can be used to obtain an estimate of cell capacity. First it should be noted that a direct estimate of the lower concentration limit, $x_f$ or $y_f$ cannot be relied upon, since a complete discharge of the battery is not expected in practice. The upper concentration limit $x_0$ or $y_0$ however can be expected to be reached during regular charging. For a given loss in capacity, the change in $x_0$ is much larger and easier to detect than the change in $y_0$. Therefore, the loss in capacity is determined by estimating the shift in $x_0$. Since the concentration corresponding to $x_f$ is not
Figure 5.12: Example of a shift in the electrode operating range from a new to an aged cell, as cyclable lithium is lost to SEI growth. In each electrode, the operating range tends to shift to a less lithiated state.

encountered in practice, $x_f$ is related to $x_0$ in a lookup table, via offline simulation of the model during a low rate, constant current discharge. For various values of $x_0$ corresponding to the upper voltage limit, the model is simulated to determine the $x_f$ corresponding to the lower voltage limit.

In this analysis, the open-loop method outlined in Chapter 4 is used to estimate the negative electrode concentration, and ultimately $x_0$. For SOC estimation, the steady-state method of applying conservation of cyclable lithium is a good approximation over a given drive cycle; however, applying the steady-state approximation for capacity estimation is somewhat counter-intuitive, since the loss of cyclable lithium is exactly what is being estimated. In other words, cyclable lithium is not expected
Figure 5.13: Visualization of the decrease in electrode compositions $x_f$ and $y_f$ resulting from a loss in lithium over a given trip. A similar diagram could be drawn for a charge to see the decrease in $x_0$ and $y_0$. In this example, assume voltage limits of 2.8–4.1 V.
Figure 5.14: Demonstration of the increased negative electrode concentration estimate error in regions where the OCV is flat. A perfect model is assumed in order to generate pseudo-data. A constant 20 mV voltage measurement bias is applied to the pseudo-data voltage. The pseudo-data is then used as the “measurement” in the EKF to estimate the negative electrode concentration.

to be conserved over the course of the aging. Therefore conceptually, the steady-state method is not as consistent with the problem of estimating loss of cyclable lithium. That said, a comparison of the open-loop and steady-state methods in Section 5.3.1 shows the capacity estimation results are similar among the two methods.

**Estimating capacity from an estimate of** $x_0$

Estimating $x_0$ and $x_f$ is not as straight forward as simply estimating the electrode concentration at a given point in time. The negative electrode OCV has regions that are relatively flat, which implies sensitivity to voltage model or sensor error. This can be demonstrated by estimating the negative electrode states for a CD cycle, where a perfect model is assumed but a 10 mV voltage measurement bias is introduced
Figure 5.15: Negative electrode OCV (measured from half cell experiment), showing regions where the OCV is flat or steep.

(Fig. 5.14). The negative electrode normalized concentration estimate is poor in regions where the OCV is flat (Fig. 5.15), but converges towards the true value in regions where the OCV is steeper. The steep OCV regions are around normalized concentrations of 0.58–0.65 and 0.22–0.35, with the latter region being the steeper of the two. However, the goal is to estimate $x_0$ (essentially the negative electrode initial condition, if the cell begins at a fully charged state), which is typically greater than 0.65, and may lie in a flat region of the OCV. This can be seen in Fig. 5.14, as the estimate is unable to accurately converge from an initial condition error at the beginning of the discharge. Therefore, to obtain an estimate of $x_0$, the strategy is to wait until the estimate is likely converged (i.e. wait until a steep OCV region is reached), and then calculate the initial state by Coulomb counting backwards in time, starting from the converged state.

Further complicating the problem, the positive electrode range, $y_0$ and $y_f$, also
changes as the cell ages. However, $y_0$ and $y_f$ estimates at a given stage of life are directly related to the estimates of $x_0$ and $x_f$, according to the following procedure:

1. Estimate $x_0$ and $x_f$.

2. Calculate $U_{0,n}$ and $U_{f,n}$ at the initial and final state via negative OCV lookup.

3. Calculate $U_{0,p} = 4.15 + U_{0,n}$ and $U_{f,p} = 2.8 + U_{f,n}$, since the upper and lower cell voltage limits remain at 4.15 V and 2.8 V throughout the battery life.

4. Calculate $y_{0,LMO}$, $y_{f,LMO}$, $y_{0,NMC}$, and $y_{f,NMC}$ via the corresponding positive OCV lookup.

5. Calculate $y_0$ and $y_f$ using (4.17) and (4.18).

The complete SOC/SOH estimation procedure is outlined in Fig. 5.16, using the open-loop method to obtain the negative electrode bulk concentration from the positive bulk concentration. Alternatively, the steady-state method could replace either the upper or lower blocks of the diagram. The observer is applied to the positive electrode states to continuously estimate the positive electrode concentrations, and the cell SOC. Then after every CD cycle, an estimate of $x_0$ and $x_f$ is obtained by waiting until either the mid SOC region or low SOC region is reached. Finally, capacity is calculated from estimates of $x_0$ and $x_f$.

Essentially there are dual observers running in parallel, one that estimates the positive electrode states to obtain cell SOC (simulating the negative electrode and liquid phase open-loop) and one that estimates the negative electrode states to obtain capacity (simulating the positive electrode and liquid phase open-loop). Clearly, the sub-models that are simulated open-loop will not be able to correct for initial condition errors. It is therefore the role of the capacity estimation algorithm to correctly update the $x_0$ and $y_0$ initial conditions so that the open-loop sub-models remain accurate.
Figure 5.16: Block diagram of the combined cell SOC/SoH estimation algorithm using the open-loop method. Cell SOC is estimated continuously from estimates of the positive electrode concentrations, while a capacity estimate is obtained once every CD cycle from an estimate of the negative electrode concentration range $x_0$ to $x_f$. 

```plaintext
\[
\begin{align*}
\text{Initialize} & \quad x_{n,k} = A_n x_{n,k-1} + B_n u_{k-1} \\
x_{n,k} &= A_n x_{n,k-1} + B_n u_{k-1} \\
\text{Negative Electrode} & \quad \xi_{n,k}, c_{n,k} \\
\text{Capacity Estimation (once per CD cycle)} & \quad \Phi_{n,k} \\
\text{Initialize} & \quad x_{p,k} = A_p x_{p,k-1} + B_p u_{k-1} \\
x_{p,k} &= A_p x_{p,k-1} + B_p u_{k-1} \\
\text{Positive Electrode} & \quad \xi_{p,k}, c_{p,k} \\
\text{Corrected} & \quad x_{p,k} = x_{p,k} - x_{e,k} \\
\text{Observer} & \quad V_{exp,k} \\
\text{Capacity Estimate} & \quad \xi_a \\
\text{Store} \quad & \quad \text{Wait until low or mid SOC region} \\
\end{align*}
\]
Table 5.1: Test conditions for the different cells tested.

<table>
<thead>
<tr>
<th>Cell</th>
<th>Test</th>
<th>Charge rate</th>
<th>Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CD to 25% SOC</td>
<td>C/3</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>CD to 25% SOC</td>
<td>3C/2</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>CD to 25% SOC</td>
<td>5C</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>CD to 35% SOC</td>
<td>C/3</td>
<td>30</td>
</tr>
<tr>
<td>5</td>
<td>CD to 45% SOC</td>
<td>C/3</td>
<td>30</td>
</tr>
</tbody>
</table>

The propagation of error resulting from this initialization procedure is a particularly important topic to study. In the capacity estimation block in Fig. 5.16, the electrode composition ranges $x_0$, $x_f$, $y_0$, and $y_f$ are used to initialize the positive and negative electrode states, which are then ultimately used to make the next electrode composition range estimates. It is conceivable that a small error in a given set of composition range estimates could contribute to additional error in the next set of estimates, and so forth, resulting in a positive feedback that would greatly amplify the initial error. This problem is investigated in Section 5.3.2, and shows that the concentration range estimates are robust enough that this positive feedback is avoided.

One advantage in the algorithm design is that the only exchange of information between the two observers comes from the update of $x_0$, $x_f$, $y_0$, and $y_f$, that are used to initialize the states. Since these parameters change only very slowly over hundreds of Ah, they have a negligible effect on the stability of the cell SOC estimation, which operates over a much faster time scale.

### 5.3.1 Validation with experimental data

To validate the combined SOC/SOH estimation algorithm, the estimation was performed on 5 different automotive pouch cells, aged under different conditions. The 5 cells differ in the final SOC that was reached in CD mode and in the charge rate, as shown in Table 5.1.
For the composite LMO-NMC cells tested, previous work used differential capacity analysis to show the dominant aging mechanism to be solid-electrolyte interface (SEI) layer growth at the negative electrode [59]. Although other mechanisms, such as loss of active material, may become significant during later stages of life [31,32] or at high temperatures, only loss of cyclable lithium is considered in this work. The SEI layer grows as a result of a lithium-consuming side reaction between the electrolyte solvent and the electrode material.

**Experimental data collection**

The same aging campaign used to validate the SOC estimates in Chapter 4, is used to validate the capacity estimates. The aging campaign was conducted using 15 Ah (nominal) automotive pouch cells with a composite LMO-NMC cathode and graphite anode, cycled under different operating conditions. The cells were cycled using the charge depleting (CD) current profile defined by the United States Advanced Battery Consortium (USABC) [1], which is representative of a plug-in hybrid vehicle (PHEV) application. The CD current profile is shown in Fig. 4.1, and was repeated until the cell reached 45, 35, or 25% SOC. Each CD cycle was followed by a constant current, constant voltage (CCCV) charge at a C/3, 3C/2, or 5C charge rate. All cells were maintained at a constant 30°C by Peltier junctions. Periodic capacity assessments were conducted at approximately every 2000-5000 Ah of throughput. The electrode OCVs (shown in Figs. 5.15 and 4.7) and parameter values from Tables 4.1-4.2 were used to calibrate the model.

An example of the open-loop model fit for the dynamic CD profile used in the experimental aging campaign is shown in Fig. 4.3. The root mean squared error over this test is 14 mV, with a maximum error of 75 mV.

It should be noted that no significant resistance rise was observed in the cells used in this work, allowing the impedance related model parameters to be treated
as constant. The change in 1 s resistance was generally less than 0.1 mΩ over the course of the aging campaign. However, for different aging conditions or electrode chemistries, resistance rise can be significant and estimation of these parameters may be required to maintain model accuracy.

Capacity estimate validation

For capacity estimation, the $x_0$ estimate is obtained by waiting for state convergence in a steep OCV region before Coulomb counting backwards to the beginning of the discharge. Therefore, a single capacity estimate is obtained after each CD cycle is complete. As noted previously, the low-SOC region from 0.22–0.35 is the preferred region to record the converged state since it corresponds to a steeper OCV than the mid-SOC region between 0.58–0.65; however, some cells are not discharged sufficiently to reach the low-SOC region. In these cases, the converged state estimate is taken in the mid-SOC region (specifically at 0.6). Otherwise if the cell does reach the low-SOC region, the estimate at mid-SOC is discarded and the converged state estimate is taken at the midway point between 0.35 and the lowest concentration reached (assuming the value is not less than 0.22). Capacity estimates for the 5 cells are shown in Figs. 5.17 - 5.21, and are compared to the measured capacities at periodic assessments. For each capacity estimate it is noted whether the estimate relied on an converged state estimate taken in the low-SOC region or mid-SOC region. Additionally, for each cell, three different estimation methods (EKF, Kalman smoother, and particle filter) are compared. The estimation results are summarized in Table 5.2.

The first thing to note is that after each capacity assessment, the estimates show some small amount of capacity recovery. This is expected since the cells would typically rest for around a day while the capacity assessments were carried out, allowing some capacity recovery.

Overall the capacity estimates are more accurate and have a lower variance for
Table 5.2: Capacity Estimation Results

<table>
<thead>
<tr>
<th>Cell</th>
<th>Observer</th>
<th>Estimate Error* Mean (Ah)</th>
<th>Estimate Error* Variance (Ah²)</th>
<th>Normalized Computation Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EKF</td>
<td>-0.08</td>
<td>0.020</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Kalman Smoother</td>
<td>0.18</td>
<td>0.023</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>Particle Filter</td>
<td>-0.04</td>
<td>0.022</td>
<td>1.81</td>
</tr>
<tr>
<td>2</td>
<td>EKF</td>
<td>-0.01</td>
<td>0.026</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Kalman Smoother</td>
<td>0.15</td>
<td>0.031</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td>Particle Filter</td>
<td>0.05</td>
<td>0.027</td>
<td>1.82</td>
</tr>
<tr>
<td>3</td>
<td>EKF</td>
<td>-0.47</td>
<td>0.012</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Kalman Smoother</td>
<td>-0.32</td>
<td>0.019</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>Particle Filter</td>
<td>-0.17</td>
<td>0.028</td>
<td>1.78</td>
</tr>
<tr>
<td>4</td>
<td>EKF</td>
<td>-0.37</td>
<td>0.081</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Kalman Smoother</td>
<td>-0.16</td>
<td>0.019</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>Particle Filter</td>
<td>-0.07</td>
<td>0.076</td>
<td>1.79</td>
</tr>
<tr>
<td>5</td>
<td>EKF</td>
<td>-0.48</td>
<td>0.073</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Kalman Smoother</td>
<td>-0.38</td>
<td>0.024</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>Particle Filter</td>
<td>-0.00</td>
<td>0.081</td>
<td>1.78</td>
</tr>
</tbody>
</table>

*Estimate error defined as the difference between each estimate and a capacity interpolated between the periodic measured capacity assessments.

Figure 5.17: Capacity estimation results using the open-loop method for Cell 1 aged, with a CD profile to 25% SOC, charged at C/3.
Figure 5.18: Capacity estimation results using the open-loop method for Cell 2, aged with a CD profile to 25% SOC, charged at 3C/2.

Figure 5.19: Capacity estimation results using the open-loop method for Cell 3, aged with a CD profile to 25% SOC, charged at 5C.
Figure 5.20: Capacity estimation results using the open-loop method for Cell 4, aged with a CD profile to 35% SOC, charged at C/3.

Figure 5.21: Capacity estimation results using the open-loop method for Cell 5, aged with a CD profile to 45% SOC, charged at C/3.
cells that discharged to the low-SOC region (Cells 1-3). This result is expected, since the estimates should be less sensitive to model and sensor errors in the low-SOC region. Also, the mid-SOC range is relatively small (only spans a normalized concentration of 0.07), so it becomes more difficult to decide when the estimate is truly converged. Essentially there is a chance of “missing” this mid-SOC region, and taking the estimate too early in the discharge (when the estimate has not yet converged) or too late (when the estimate is starting to diverge again). At BOL, cells 4 and 5 are not discharged sufficiently to reach the low-SOC region, so the initial estimates for these cells are taken in the mid-SOC region. However, as the estimated $x_0$ decreases over time, the end of the CD cycles begin to cross the threshold into the low-SOC region, at which point the algorithm switches to estimate in the low-SOC region. An important point here is that although the estimated concentration may cross into the low-SOC region, triggering the switch, the actual cell may not have done so. This is most likely the case for cell 5, which is only discharged to 45% SOC. Right before the switch to estimating at low-SOC is made near the end of the test, the estimated capacities happen to be lower than the measured capacity (at least for the EKF and smoother). This means that the switch is made prematurely, resulting in additional inaccuracies in the estimate after the switch is made. That said, the EKF and smoother eventually switch back to estimating in the mid-SOC region, tending to oscillate between the two regions. These problems illustrate the difficulty in relying on an imperfect estimate to determine if thresholds have been crossed.

In comparing the three observer types, there are not large differences in the estimate error mean or variance among the observers. The particle filter tends to produce slightly more accurate but noisier estimates than the other observers, particularly when estimating in the mid-SOC region. Overall, the EKF performs well without the additional computational effort required for the smoother and particle
Comparison between steady-state and open-loop methods

A direct comparison of the open-loop and steady-state methods (for the EKF only) is shown in Figs. 5.22-5.26.

Both the open-loop and steady-state methods produce very similar results using the EKF. The only major difference is for cell 5, where the steady-state method avoids the premature switching to estimate at low-SOC near the end of the cycling; though this is most likely not a general result. Overall the similar results produced by each method suggests that both open-loop and steady-state methods are likely viable. Furthermore, the similarity among methods indicates that the open-loop method is not overly affected by initialization errors that are generally detrimental to open-loop estimates. The open-loop method may be more useful in predicting available power,
Figure 5.23: Capacity estimation results for Cell 2, comparing the open-loop and steady-state methods.

Figure 5.24: Capacity estimation results for Cell 3, comparing the open-loop and steady-state methods.
Figure 5.25: Capacity estimation results for Cell 4, comparing the open-loop and steady-state methods.

Figure 5.26: Capacity estimation results for Cell 5, comparing the open-loop and steady-state methods.
since it includes the diffusion dynamics and improves the model for high current/low SOC conditions. However, the steady-state method shows some improved robustness to errors in the composition ranges for SOC estimation. It is also possible to use, say, the steady-state method for continuous cell SOC estimation and the open-loop method for periodic capacity estimation.

**Maintaining SOC accuracy**

In addition to estimating cell capacity, another important objective in estimating the changes in model parameters corresponding to the dominant degradation mode within a cell (e.g. \( x_0 \) as it relates to loss of cyclable Li) is to maintain an accurate model over time, thereby maintaining an accurate SOC estimate over time. Figure 5.27 shows how the accuracy of the SOC estimate changes over time for cell 2 (the other cells showed a similar result). The estimated \( x_0 \) is used to update the model after each CD cycle, and as described previously, the SOC estimation is done using the EKF on the positive electrode sub-model. The capacity used to calculate the measured SOC by Coulomb counting is obtained by interpolating between each periodic capacity assessment.

The SOC estimate error over a given CD-CCCV profile remains fairly constant throughout the life of the cell, indicating that the model is being updated correctly with estimates of \( x_0 \). Note that the outliers in the plot are instances where the profile had two CCCV charge portions, over which the model is less accurate.

**5.3.2 Quantifying SOC/SOH Estimate Error**

In this section, an attempt is made to quantify the SOC and capacity estimate error. For this analysis, the EKF is used as the observer, though a similar analysis could be done using any nonlinear observer. Also, this analysis assumes that the open-loop
method is used, where the EKF is only applied to the states of one electrode.

Before discussing the sources of estimate error, it should be noted that the SOC and capacity measurements (ultimately these are measurements of current) also have some associated error, depending on the accuracy of the power supply used to cycle the cells. The Amrel power supply lists the measured current accuracy at 0.3% ± 20 mA. This can be used to establish error bars on the SOC and capacity measurements, which are based on integration of the current measurements.

The sources of SOC and capacity estimate error can be broken down into 3 categories: errors due to process and sensor noise, model error, and errors due to incorrect initial composition range. State estimate errors due to process and sensor noise are captured by the estimate covariance matrix, which is estimated by the EKF. Model errors result in a deviation between the predicted and measured output, resulting in a “false” output residual that causes an unnecessary state correction. Finally, errors
in the electrode composition ranges, \( x_0, x_f, y_0, \) and \( y_f \), which are used to initialize the electrode states, will contribute to state error in the electrode that is simulated open-loop and cannot correct for an initial condition error. The propagation of these 3 error types for the SOC and capacity estimates is shown in Fig. 5.28.

First, the error in the capacity \( (x_0) \) estimates is discussed, followed by a discussion of how the capacity estimate error goes on to affect the SOC estimate error.

**Capacity estimate error**

Recall that to estimate capacity using the open-loop method, only the negative electrode states are estimated. At an appropriate time step \( k \) (either at mid-SOC or low-SOC), an estimate of \( x_0 \) is made by Coulomb counting backwards from the state estimate \( \hat{x}_k \). Each error type is first treated separately.
1. Error due to noise

The estimate error due to the presence of process and sensor noise is directly estimated by the EKF, via the estimate covariance matrix, \( P \). This covariance matrix contains variances of the state estimate error on its diagonal, and the covariances between states on the off-diagonals. To isolate the error associated with each state, a state transformation can be applied to diagonalize \( P \) [91]:

\[
\tilde{x} = M^T x \quad (5.8)
\]

\[
\tilde{P} = M^T P M \quad (5.9)
\]

where the columns of the transform matrix, \( M \), are the eigenvectors of \( P \). This results in a transformed covariance matrix that consists of only variances on its diagonal, in other words:

\[
\tilde{P} = \begin{bmatrix}
\sigma_{\tilde{x}_1, \tilde{x}_1}^2 & 0 \\
0 & \ddots \\
0 & 0 & \sigma_{\tilde{x}_n, \tilde{x}_n}^2
\end{bmatrix} \quad (5.10)
\]

The error for each state can be bounded at the \( 2\sigma \) level, so that:

\[
\tilde{x} = \tilde{x} \pm \text{diag} \left( 2\sqrt{\tilde{P}} \right) \quad (5.11)
\]

where \( \text{diag} \) indicates taking the elements along the diagonal. The bound on the transformed state can be converted back to a bound on the original state by changing coordinates again:

\[
x = x \pm e_x = x \pm M \text{diag} \left( 2\sqrt{P} \right) \quad (5.12)
\]
Figure 5.29: Absolute value of the difference between predicted and measured cell voltage, over the course of a CD cycle.

The states are related to SOC through a simple linear combination, so to obtain a bound on the bulk SOC of a given electrode particle, the error in each state compounds to give:

$$\bar{SOC}_i = \bar{SOC}_i \pm \sqrt{\left( \frac{a_{0,i}}{c_{max,i}} e_{x_1} \right)^2 + \left( \frac{a_{1,i}}{c_{max,i}} e_{x_2} \right)^2 + \left( \frac{a_{2,i}}{c_{max,i}} e_{x_3} \right)^2} \quad (5.13)$$

Using the CD-CCCV current profile described previously, the SOC error bounds compute to $< \pm 0.1\%$ over the course of the cycle; however, this result is highly dependent on the choice of covariance matrices, $Q$ and $R$. Clearly, this is not a sufficient error bound in practice, as the presence of model error and initial condition error (in the electrode that is simulated open-loop) will also negatively affect the state estimate.

2. Model error
The presence of model error, i.e., a difference in the predicted voltage output compared to the measured voltage, can negatively affect the state estimate. Model error can be attributed either to an error in the assumed state dynamics or in the assumed output equation. A sensor bias, while not technically a model error, would have a similar effect on the voltage residual, so it can be grouped into this category as well. The difference in predicted and measured cell voltage is shown in Fig. 5.29, over the course of a CD cycle down to 25% SOC. The root-mean-square (RMS) error over this cycle is 14 mV, with a maximum error of 75 mV. To understand how this model error affects the state estimate of the negative electrode, consider the following case shown in Fig. 5.30, where a perfect model is assumed but a constant 14 mV voltage bias is introduced in either direction. ¹ Again, note that the estimate error is much larger during regions where the negative OCV is very flat. However, the algorithm for estimating \( x_0 \), and therefore capacity, relies only on the state estimate during the regions where the negative OCV is steep (either at mid-SOC or low-SOC). In these regions, the \( SOC_n \) is less affected by the model error. Once the state converges towards the true value in a steep OCV region, the error in the estimated \( SOC_n \) is generally < 5%. In the worst case, the maximum \( SOC_n \) estimate error is 16% in the mid-SOC region and 8% in the low-SOC region.

3. Error due to incorrect initialization

In the open-loop method for estimating the negative electrode states (and also in the steady-state method), the composition ranges \( x_0, x_f, y_0, \) and \( y_f \) are used to

¹Rather than use experimental data, a perfect model is used here to generate pseudo-data so that the “true” \( SOC_n \) is known. Applying, say, the maximum error of 75 mV as a bias to the pseudo-data rather than the RMS error would give a much more conservative \( SOC_n \) error bound, but a less useful one, as this is not realistic. Alternatively, the exact error profile seen in Fig. 5.29 could be imparted on the pseudo-voltage data; however, this error is dependent on the particular current profile. Simply applying the constant 14 mV RMS error is a more generalizable approach.
initialize each electrode, via the procedure outlined in Section 4.2.4. Since in this case the positive electrode is simulated open-loop, any errors in the composition ranges will negatively affect the open-loop prediction of the positive electrode voltage, as these errors cannot be corrected through measurement feedback. An error in the predicted positive electrode voltage will, in turn, impose an error in the closed-loop estimates of the negative electrode states (see Fig. 5.16), and an error in the next $x_0$ estimate. This propagation of $x_0$ estimate error is shown in Fig. 5.28. It is conceivable that a small error in a given $x_0$ estimate could contribute to additional error in the next $x_0$ estimate, and so forth, resulting in a positive feedback that would greatly amplify the initial error. To test whether this positive feedback occurs, consider again an assumed perfect model to generate pseudo-data over a CD cycle. However in this case, an error in the initial $x_0$ is imposed. This initial $x_0$ error will cause errors in the other
Figure 5.31: Error in the $x_0$ estimate during a CD cycle, in response to an error in the initial assumed $x_0$. Here, a perfect model is assumed.

composition range parameters ($x_f$, $y_0$, and $y_f$), and will adversely affect the negative electrode state estimate and subsequent $x_0$ estimate.

The error in the $x_0$ estimate as a function of various initial $x_0$ errors is shown in Fig. 5.31. The solid black lines drawn on the plot represent the lines $y = x$ and $y = -x$. The location of the data points in relation to these lines determines whether or not the algorithm is stable. A stable algorithm requires that the $x_0$ estimate error is less than the initial $x_0$ error. In other words, the $x_0$ error should decrease with each iteration. Regions of stability (where the $x_0$ estimate error is less than the initial $x_0$ error) and instability (where the $x_0$ estimate error is greater than the initial $x_0$ error) are marked on the plot. If, for example, a point were to lie in an unstable region, the $x_0$ error would grow (shifting the point to the left or right for the next iteration). As the point shifts to the right or left with each iteration, it is possible for it to eventually enter a stable region. At
this point, the \( x_0 \) error would settle around some nominal value corresponding to a point near or at one of the black lines.

Clearly, the algorithm is insensitive enough to initial \( x_0 \) errors that almost all of the points lie in stable regions, so that the \( x_0 \) error decreases with each iteration. The only exception is for a near zero initial \( x_0 \) error, which would quickly stabilize at some small value (< 0.1%). This analysis shows that an \( x_0 \) estimate can recover from an imposed initial \( x_0 \) error (at least with an assumed perfect model). One reason for this robustness to initial \( x_0 \) errors is that the effect of an initial \( x_0 \) error on \( y_0 \) (and the resulting open-loop positive voltage) is minor, due to the flat negative electrode OCV near the upper voltage limit. Essentially, a large error in \( x_0 \) produces a small error in the initial \( U_n \), which in turn produces a small error in the initial \( U_p \) and \( y_0 \).

So far, each source of capacity estimate error has been analyzed individually. This has revealed a negligible state estimate error due to process and sensor noise (< 0.1%), a potentially significant state estimate error due to model error (generally < 5%, but up to 16%), and a negligible \( x_0 \) estimate error due to incorrect initial composition ranges (< 0.1%). However, there is no guarantee that the net effect of each error type will be additive. Therefore, the three sources of error are now applied simultaneously. The same 14 mV voltage bias is applied (in both directions) to the model-generated pseudo-data, and an initial \( x_0 \) error is imposed. The result is shown in Fig. 5.32. Rather than plot a single point for each initial \( x_0 \) error, the shaded areas represent all possible results for estimates taken at any point within the mid-SOC or low-SOC regions. Again these regions are \( SOC_n = 0.58 - 0.65 \) and \( SOC_n = 0.22 - 0.35 \), respectively.

By looking at all possible \( x_0 \) estimates made in each region, the “worst-case” scenario can be singled out and used to bound the \( x_0 \) estimates. As mentioned
Figure 5.32: Error in the $x_0$ estimate during a CD cycle, in response to an error in the initial assumed $x_0$. Here a 14 mV voltage bias is applied to the model-generated pseudo-data. The shaded areas represent all possible results for estimates taken at any point within the mid-SOC or low-SOC regions.

Previously, points that lie in an unstable region will tend to shift to the right or left with each iteration. Upon crossing one of the black lines and entering a stable region, the $x_0$ estimate error will settle around some nominal value near the black line. These “settling points” are marked on the plot for the worst case scenarios at mid-SOC and low-SOC. The $x_0$ estimate error corresponding to these settling points represents the worst possible $x_0$ error in response to both model error and an initial $x_0$ error. Taking the settling point with the largest $x_0$ estimate error value gives a maximum $x_0$ error of 7% for the estimates taken at low-SOC and 16% for the estimates taken at mid-SOC. Interestingly, the cumulative effect of both model error and initialization error is approximately equal to the sum of the individual effects.

The maximum $x_0$ error can be converted into a maximum capacity error, via Eq. (5.7), resulting in a capacity error bound of ± 0.53 Ah for estimates taken in the
Figure 5.33: Capacity estimation results for cell 1, with dashed lines indicated the computed error bounds of $\pm 0.53$ Ah, applied to the EKF estimates.

low-SOC region and $\pm 2.0$ Ah for estimates taken in the mid-SOC region. These bounds are applied to the EKF estimation results in Figs. 5.33-5.34 for cells 1 and 5, where the measurement error bars are also shown. The measured capacities all fall within these bounds. Also, the estimates using the Kalman smoother and particle filter also fall within the error bounds. The bounds are conservative in some respects, since they account for the worst case effects from incorrect initialization, and consider a constant voltage bias. In most cases, model error will not be reflected as a constant bias, but rather some dynamic trajectory that will tend to average out over a given cycle. One exception could be in the case of a voltage sensor error, where a constant bias of up to 10 mV could be encountered.

The main result of this analysis is that by far the limiting factor in the capacity estimate accuracy is model error. This is an important result, as model error is something that can be improved with better knowledge of the electrode and electrolyte.
Figure 5.34: Capacity estimation results for cell 5, with dashed lines indicated the computed error bounds of ± 2.0 Ah, applied to the EKF estimates.

properties. This is better than if, say, the limiting factor were due to inaccurate initialization or the open-loop approximation itself, which are intrinsic to the algorithm and cannot necessarily be improved without a reference electrode. Many of the electrochemical parameters used in the model were taken from literature or fit to match experimental data, and may not represent the specific material properties used in these automotive cells. For example, the individual OCVs of LMO and NMC and their volume fraction split were not available from the cell manufacturer. The OCVs in particular are one of the most important components of the model that determine its accuracy, and these had to be taken from literature. In short, there is reason to believe that the capacity estimate results could be significantly improved with better knowledge of the model parameters.
SOC estimate error

After putting reasonable bounds on the \( x_0 \) and capacity estimate error, it is possible to determine how this \( x_0 \) error, combined with model error, affect the SOC estimate. Again using the open loop method, a perfect model is assumed with a constant 14 mV voltage bias and a 7% or 16% error in \( x_0 \) added (corresponding to estimates taken at low-SOC and mid-SOC, respectively). The voltage bias and \( x_0 \) error are applied in both directions, first individually then together. The range of outcomes is shown in Fig. 5.35.

The effect of the constant 14 mV voltage bias is at maximum a 3% error in estimated SOC. The 7% error in \( x_0 \) for \( x_0 \) estimates taken at low-SOC results in a maximum 2% SOC estimate error with a perfect model and a maximum 4% error with the 14 mV voltage bias applied as well. The 16% error in \( x_0 \) for \( x_0 \) estimates taken at mid-SOC results in a maximum 8% SOC estimate error with a perfect model and a maximum 12% error with the added 14 mV voltage bias. Clearly in this case, both model error and \( x_0 \) error contribute to the overall SOC estimate error. Again, with improved model parameterization, both of these effects would diminish, resulting in a more accurate SOC estimate.

Applying the algorithm to experimental data, the SOC estimate bounds are shown in Fig. 5.36, along with the SOC measurement error, for one CD cycle taken at BOL. In this plot, only the estimate error bound associated with the \( x_0 \) estimate taken at low-SOC is shown. The measured SOC bounds grow as the current measurement is integrated over time, and become on the same order as the estimate error near the end of the cycle (3-4%). The measured SOC lies well within the uncertainty in the estimate.

One additional source of estimate error that was not considered in this analysis is the effect of current sensor error on the state estimates. Including this effect would
Figure 5.35: Cell SOC error in response to a.) a ± 14 mV voltage bias, b.) a 7% or 16% error in $x_0$, corresponding to $x_0$ estimates taken at low-SOC or mid-SOC, respectively, and c.) both voltage bias and $x_0$ error.
further widen the error bounds on both the capacity and SOC estimates. That said, the SOC estimate error resulting from a current bias of $0.3\% \pm 20\,\text{mA}$ is less than 1% at the end of a given CD-CCCVC profile, as the voltage feedback provides some robustness to this input sensor error.

## 5.4 Simultaneous Estimation of Degradation Mechanisms

In the previous sections, algorithms are designed to estimate the parameters associated with loss of active material and loss of cyclable lithium, with the assumption that each mechanism is acting alone. This allows one electrode to be simulated open-loop, since the degradation only affects the parameters of one electrode (the positive elec-
trode in the case of estimating $\epsilon_i$ and the negative electrode in the case of estimating $x_0$). In this section, an attempt is made to consider the case where both loss of active material and loss of cyclable lithium are occurring simultaneously. In [32], it is shown that for the LMO-NMC composite cells, loss of cyclable lithium occurs throughout the cell life but loss of active material can accelerate the aging at later stages of life. Performing a differential capacity analysis on our own commercial LMO-NMC cells did not reveal any loss of active material, though the aging campaign did not continue past around 10% capacity loss for most cells. Furthermore, most cells were not cycled at high temperatures, which made them more resistant to active material loss (Mn dissolution in particular). Therefore, instead of using the experimental aging campaign to validate any combined active material/cyclable lithium estimation algorithm, the validation is done in simulation only. However, model error may have a significant effect on the estimate accuracy (as shown in Section 5.3.2), so it is important to not simply rely on a “perfect model” assumption to generate pseudo-data.

In this section, a combined active material/cyclable lithium estimation algorithm using an EKF is tested under various scenarios, including the assumption of a perfect model, the assumption of a dynamic model error, and the assumption of a voltage sensor bias. In light of the observability problems seen when attempting to estimate the states of both electrodes, it is also shown how the estimate accuracy can improve if a reference electrode is available.
5.4.1 Combined active material/cyclable lithium estimation algorithm

In order to estimate the parameters $\epsilon_i$ and $x_0$, associated with loss of active material and loss of cyclable lithium, respectively, the complete state space model is as follows:

\[
\dot{x}(t) = \begin{cases} 
  x_2(t) \\
  x_3(t) \\
  -\frac{1}{b_{3,LMO}} x_2(t) - \frac{b_{2,LMO}}{b_{3,LMO}} x_3(t) + \frac{R_{LMO}}{3b_{3,LMO} F_{ALP} \tau(t)} \beta_{LMO}(t) u(t) \\
  x_5(t) \\
  x_6(t) \\
  -\frac{1}{b_{3,NMC}} x_5(t) - \frac{b_{2,NMC}}{b_{3,NMC}} x_6(t) + \frac{R_{NMC}}{3b_{3,NMC} F_{ALP} x_8(t)} \beta_{NMC}(t) u(t) \\
  0 \\
  0 \\
  x_{10}(t) \\
  x_{11}(t) \\
  -\frac{1}{b_{3,n}} x_{10}(t) - \frac{b_{2,n}}{b_{3,n}} x_{11}(t) - \frac{R_n}{3b_{3,n} F_{ALp} \epsilon_n} u(t)
\end{cases}
\]

\[
V(t) = U_i(c_{s,i}(t)) - \eta(c_{s,i}(t), \epsilon_i(t), \beta_i(t) u(t)) - R_{c,i} \beta_i(t) u(t) \\
- U_n(c_{s,n}(t)) - \eta(c_{s,n}(t), u(t)) - R_{cell} u(t), \quad i = LMO, NMC
\]
the liquid phase potential drop, with the assumption that it can always be simulated open-loop and added to the cell voltage measurement.

After each CD cycle, estimates of $\epsilon_{LMO}$, $\epsilon_{NMC}$, and $x_0$ are obtained, and the model is updated accordingly. The $\epsilon_i$ estimates obtained at the end of each CD cycle are carried over to initialize $\epsilon_i$ for the next cycle. For the $x_0$ parameter, the same process outlined in Section 5.3 is used, where first an estimate of the negative electrode concentration in the low SOC region is obtained, then Coulomb counting is used to back-track to the beginning of the CD cycle in order to estimate $x_0$. This $x_0$ estimate is then used to initialize the electrodes in the next cycle.

Voltage pseudo-data is generated by simulating the single particle model over 200 consecutive CD cycles, where both $\epsilon_{LMO}$ and $x_0$ are decreased linearly to represent loss of LMO active material and loss of cyclable lithium. Noise is also added to the voltage pseudo-data.

As explained previously, 3 scenarios are considered: a perfect model (plus voltage sensor noise), a dynamic model error, and a voltage sensor bias. The dynamic model error is generated by using the difference between the open-loop model prediction and the experimental voltage, measured from one of the composite LMO-NMC pouch cells over a CD cycle. The resulting dynamic model error is shown in Fig. 5.37. The comparison to obtain the model error is done using data at BOL, so this represents the nominal model error that can be expected when the model parameters are first initialized. This dynamic model error is then super-imposed on to the voltage pseudo-data for each CD cycle. Finally in the 3rd scenario, a voltage sensor bias of $\pm 10\text{mV}$ is added to the pseudo-data to represent the limited accuracy of an onboard voltage sensor.

The estimation results are shown in Figs. 5.38-5.39, under each scenario, where a 20% initial condition error is imposed on the initial $x_0$, $\epsilon_{LMO}$, and $\epsilon_{NMC}$ estimates.

In the case of a perfect model, the estimates are able to converge from the initial
Figure 5.37: Model error over the course of the CD profile, defined as the difference between the measured cell voltage and the model prediction. This dynamic error profile is then imposed on the voltage pseudo-data to replicate experimental data. Condition error and track the true values with a small bias. However, introduction of the dynamic model error and sensor bias, result in very inaccurate estimates. Particularly in the case of adding the dynamic model error, the estimates are highly inaccurate. Clearly, the poor observability of the complete model (made worse by the addition of the volume fraction states) allows for very little robustness to model or sensor errors using the EKF. This is not necessarily a general result for all observers. For example, [76] showed that using the UKF for estimation of cyclable lithium and active material gives a better result than using the EKF, under the assumption of a perfect model (plus zero-mean noise). However, it is unclear if this result has any robustness in the presence of model or sensor error, as the authors did not test this. Interestingly, the result in Fig. 5.39 using the perfect model is much better than the EKF result in [76], in which the estimates essentially become unstable even with
Figure 5.38: $x_0$ estimates under scenarios of a perfect model, dynamic model error, and voltage sensor bias.

It should be noted that the estimate convergence is particularly sensitive to the tuning of the noise covariances, and the instability seen with a perfect model in [76] can be replicated by tuning the filter differently.

**Adding a reference electrode**

The poor robustness of the estimation algorithm to model and sensor error is ultimately due to the lack of suitable measurements. Essentially, there is a limit to how much information about the internal states that can be realistically obtained using only current and voltage measurements. However, the use of a reference electrode can greatly improve the observability of the model states. A reference electrode could be used to obtain a direct measurement of each electrode voltage, relative to the reference. Using a pure lithium reference electrode, for example, would minimally affect the cell voltage dynamics, since redox reactions at a pure lithium electrode are very
Figure 5.39: Positive electrode active material volume fraction estimates under scenarios of a perfect model, dynamic model error, and voltage sensor bias.

Fast [12]. From an observability perspective, a reference electrode allows the model to be split into two sub-systems, each with its own measurement. The resulting observability matrix for each sub-system would be identical to the observability matrix using the open-loop method, as shown in Section 4.3.3, resulting in a significantly lower condition number. Of course, rather than relying on an open-loop estimate of the opposite electrode, a reference electrode would allow the states of both electrodes to be estimated closed-loop, while still maintaining the improved observability seen when a direct electrode voltage measurement is available.

To test how the estimate accuracy might be improved with a reference electrode,
the combined active material/cyclable lithium estimation algorithm is now implemented with the assumption that a direct measure of each electrode voltage is available. The same pseudo-data is used, under the scenarios of a perfect model, dynamic model error, and sensor bias. The model can be broken into two sub-models, one for each electrode. The estimate of cyclable lithium loss, i.e. a decrease in $x_0$, is obtained from the negative electrode sub-model, while the estimates of active material loss are obtained from the positive electrode sub-model. Figures 5.40-5.41 show the $x_0$ and volume fraction estimation results, with the assumption of a reference electrode. Again, a 20% initial condition error is imposed on the initial $x_0$, $\epsilon_{LMO}$, and $\epsilon_{NMC}$ estimates.

As expected, the assumption of a perfect model allows the $x_0$ and volume fraction estimates to converge to the true values accurately. The imposed dynamic model error and sensor bias result in biased, but stable parameter estimates. Considering the
Figure 5.41: Positive electrode active material volume fraction estimates under scenarios of a perfect model, dynamic model error, and voltage sensor bias, assuming a reference electrode is available.

worst case errors of +0.073, +0.062, and -0.056 for $x_0$, $\epsilon_{LMO}$, and $\epsilon_{NMC}$, respectively, this results in a maximum cell capacity estimate error of 6.4%. This represents a significant improvement compared to the case without a reference electrode, in which the addition of the dynamic model error or 10 mV sensor bias can lead to highly inaccurate estimates.

5.5 Summary and Future Work

State of health, while not uniquely defined, represents the amount of degradation that a cell has experienced over its life, contributing to capacity and power fade.
In the context of electrochemical models, SOH estimation involves estimating model parameters that correspond to specific degradation mechanisms like loss of active material and loss of cyclable lithium. An algorithm to estimate loss of active material in a composite positive electrode is applied to data collected from fabricated coin cells that were repeatedly cycled to induce aging. These estimates of active material are able to correctly predict the overall measured capacity loss as well as correctly predict the relative loss of each composite material.

Additionally, an algorithm to estimate loss of cyclable lithium is applied to data collected from commercial cells aged under various conditions, where the primary degradation mechanism is found to be SEI growth. Loss of cyclable lithium corresponds to a shift in the composition ranges of each electrode, towards a more delithiated state. An estimate of how these composition ranges change gives an estimate of the battery capacity. The algorithm takes advantage of the fact that model parameters change very slowly as the battery ages, allowing the possibility to wait until ideal conditions before making an estimate. Specifically, flat regions of the graphite electrode OCV produce inaccurate state estimates and these regions can be avoided. Capacity estimates are validated experimentally by comparing with measured capacities obtained periodically throughout the aging campaign. In general, capacity estimates for cells that were discharged to a low SOC show improved accuracy over cells that were only discharged to a mid SOC. In addition, an error analysis of the parameter estimation algorithm reveals that the primary source of estimate error is due to model error, something that can be improved with better knowledge of model parameters.

An attempt is made to estimate the parameters associated with both loss of active material and loss of cyclable lithium, if both mechanisms occur simultaneously. This necessitates using the complete battery model without using the open-loop or steady-state approximations, resulting in a model that is shown to be poorly observable.
Applying the EKF to estimate loss of active material and lithium in a simulated composite LMO-NMC cell, shows that the estimates can converge to the true values with the assumption of a perfect model, but become highly inaccurate in the presence of model error or a voltage sensor bias. However, by using a reference electrode, the estimates show much better accuracy, even under model and sensor error.

Future work in this area could focus on applying different nonlinear observer types in order to improve robustness to model error. In particular the Kalman smoother, particle filter, or UKF could be investigated; although an emphasis should still be placed on keeping the algorithm computationally efficient. Experimental validation of the estimation algorithms with a reference electrode would provide further quantification of the estimate accuracy under real world conditions, and could be a strong motivator to investigate the feasibility of embedding one or more reference electrodes in an actual automotive pack. Second, the estimation of resistance related parameters that might also change as the cell ages could be investigated. From an observability perspective, the feasibility of estimating even more model parameters simultaneously is questionable; however, it certainly may be possible to do so with careful selection of when to make an estimate, such as immediately following a current pulse, or with the use of a reference electrode. Third, validation of the SOH estimation algorithms at different operating temperatures is a critical step towards determining their feasibility. This would involve scheduling certain model parameters like reaction rate constants, conductivity, and diffusion coefficients based on the measured cell temperature. An important part of this validation would be to ensure that voltage changes within a given drive cycle due to aging could be distinguished by the algorithm from voltage changes due to changes in operating temperature. An accurate scheduling of model parameters based on temperature would be critical for the aging effects to be distinguishable. Finally, the model structure used in this work could be applied towards prognosis problems, in order to develop a “remaining useful life” indicator. In
this case, the evolution of model parameters related to aging could be projected into the future, using a semi-empirical or physical model of the corresponding degradation mechanism.
Chapter 6

Summary

In this dissertation, the application of reduced order electrochemical models towards SOC and SOH estimation problems is investigated. A simplified model of a composite electrode is obtained by applying various model order reduction techniques, while still preserving model parameters that have physical meaning. This model is then used to estimate parameters associated with SOC and SOH. In the case of SOH estimation, loss of active material and loss of cyclable lithium are the two main degradation mechanisms considered. The parameter estimation algorithms are validated in simulation and experimentally, through an aging campaign with composite LMO-NMC cells.

In summary, electrochemical models have the potential to provide more meaningful information about the inner workings of the battery. These models can be used to estimate parameters associated with specific degradation mechanisms, and facilitate more accurate and robust diagnostics and prognostics of the battery. The analysis and validation of SOC and SOH estimation algorithms done in this dissertation show both the advantages and challenges of this approach. While the results presented here are promising, the biggest limitation is the lack of a reference electrode that would help distinguish degradation in the anode and cathode. Having this additional measurement would truly allow the electrochemical model-based approach to obtain
its full potential.
Bibliography


[34] A.D. Filippi, S. Stockar, S. Onori, M. Canova, and Y. Guezenec. Model-based life estimation of li-ion batteries in phevs using large scale vehicle simulations:


243


Appendix A

Observability Calculations

A.1 Observability of the 2-Electrode Single Particle Model

The state dynamics of the complete single particle model with a particle for each electrode can be written as:

\[
\dot{x} = f(x, u) = \begin{bmatrix}
    x_2 \\
    x_3 \\
    -\frac{1}{b_{3,p}}x_2 - \frac{b_{2,p}}{b_{3,p}}x_3 + \gamma_p u \\
    x_5 \\
    x_6 \\
    -\frac{1}{b_{3,n}}x_5 - \frac{b_{2,n}}{b_{3,n}}x_6 - \gamma_n u
\end{bmatrix}
\]  

(A.1)

\[
\begin{bmatrix}
    c_{s,p} \\
    \bar{c}_p \\
    c_{s,n} \\
    \bar{c}_n
\end{bmatrix} = \begin{bmatrix}
    a_{0,p} & a_{1,p} & a_{2,p} & 0 & 0 & 0 \\
    a_{0,p}b_{2,p} & a_{0,p}b_{3,p} & 0 & 0 & 0 \\
    0 & 0 & 0 & a_{0,n} & a_{1,n} & a_{2,n} \\
    0 & 0 & 0 & a_{0,n}b_{2,n} & a_{0,n}b_{3,n}
\end{bmatrix}
\begin{bmatrix}
    x
\end{bmatrix}
\]
\[ V = g(x, u) = (s_{3,p}c_{s,p}^3 + s_{2,p}c_{s,p}^2 + s_{1,p}c_{s,p} + s_{0,p}) \]
\[- \eta_p(c_{s,p}, u) - R_{c,p}u \]
\[- (s_{3,n}c_{s,n}^3 + s_{2,n}c_{s,n}^2 + s_{1,n}c_{s,n} + s_{0,n}) \]
\[- \eta_n(c_{s,n}, u) - R_{c,n}u \] (A.2)

where \(\gamma_i = \frac{R_i}{3b_i,F_{AL,i,c_{am,i}}}, s_{j,i} \) are the OCV cubic spline coefficients, \(c_{s,i} \) and \(\bar{c}_i \) are linear functions of the states, and \(a_{j,i} \) and \(b_{j,i} \) result from the 3rd order Padé approximation.

The liquid dynamics are neglected here, with the assumption that they can be modeled open-loop. The 0th Lie derivative \(L_0^f(g)\) is simply equal to the output:

\[ L_0^f(g) = g(x, u) \] (A.3)

The first Lie derivative is equal to:

\[ L_1^f(g) = \frac{\partial g}{\partial x_1} f_1(x, u) + \ldots + \frac{\partial g}{\partial x_6} f_6(x, u) + \frac{\partial g}{\partial u} \dot{u} \]. The calculation of each term is as follows:

\[ \frac{\partial g}{\partial x_1} f_1(x, u) = \left(3s_{0,p}c_{s,p}^2a_{0,p} + 2s_{1,p}c_{s,p}a_{0,p} + s_{2,p}a_{0,p} + \frac{\partial \eta_p}{\partial x_1}\right)x_2 \]
\[ \frac{\partial g}{\partial x_2} f_2(x, u) = \left(3s_{0,p}c_{s,p}^2a_{1,p} + 2s_{1,p}c_{s,p}a_{1,p} + s_{2,p}a_{1,p} + \frac{\partial \eta_p}{\partial x_2}\right)x_3 \]
\[ \frac{\partial g}{\partial x_3} f_3(x, u) = \left(3s_{0,p}c_{s,p}^2a_{2,p} + 2s_{1,p}c_{s,p}a_{2,p} + s_{2,p}a_{2,p} + \frac{\partial \eta_p}{\partial x_3}\right) \cdot \left(-\frac{1}{b_{3,p}}x_2 - \frac{b_{2,p}}{b_{3,p}}x_3 + \gamma_p u\right) \]
\[ \frac{\partial g}{\partial x_4} f_4(x, u) = -\left(3s_{0,n}c_{s,n}a_{0,n} + 2s_{1,n}c_{s,n}a_{0,n} + s_{2,n}a_{0,n} + \frac{\partial \eta_n}{\partial x_4}\right)x_5 \] (A.4)
\[ \frac{\partial g}{\partial x_5} f_5(x, u) = -\left(3s_{0,n}c_{s,n}a_{1,n} + 2s_{1,n}c_{s,n}a_{1,n} + s_{2,n}a_{1,n} + \frac{\partial \eta_n}{\partial x_5}\right)x_6 \]
\[ \frac{\partial g}{\partial x_6} f_6(x, u) = -\left(3s_{0,n}c_{s,n}a_{2,n} + 2s_{1,n}c_{s,n}a_{2,n} + s_{2,n}a_{2,n} + \frac{\partial \eta_n}{\partial x_6}\right) \cdot \left(-\frac{1}{b_{3,n}}x_5 - \frac{b_{2,n}}{b_{3,n}}x_6 - \gamma_n u\right) \]
\[ \frac{\partial g}{\partial u} \dot{u} = \left(\frac{\partial \eta_p}{\partial u} - \frac{\partial \eta_n}{\partial u} - R_{c,p} - R_{c,n}\right)\dot{u} \]
Here the partial derivatives of $\eta_i$ have not been expanded but are straightforward to calculate. The remaining Lie derivatives won’t be expanded, but follow from differentiating the previous Lie derivative in time. So the second Lie derivative is:

$$L^2_f(g) = \frac{\partial L^1_f(g)}{\partial x_1} f_1(x, u) + \ldots + \frac{\partial L^1_f(g)}{\partial x_6} f_6(x, u) + \frac{\partial L^1_f(g)}{\partial u} \dot{u} + \frac{\partial L^1_f(g)}{\partial \dot{u}} \ddot{u},$$
and so on. Note that for the $n$th Lie derivative, there is a dependence on the input and the first $n$ time derivatives of the input. It is therefore possible for certain systems that both a non-zero input and an input that has a certain number of non-zero derivatives are required for the system to be observable. Fortunately, this is not the case for the single particle model.

Finally, the observability matrix is obtained by taking the Jacobian of the vector of Lie derivatives:

$$O' = \begin{bmatrix}
\frac{\partial L^0_f(g)}{\partial x_1} & \ldots & \frac{\partial L^0_f(g)}{\partial x_6} \\
\vdots & \ddots & \vdots \\
\frac{\partial L^5_f(g)}{\partial x_1} & \ldots & \frac{\partial L^5_f(g)}{\partial x_6}
\end{bmatrix} \quad (A.5)$$

The calculation of the observability matrix using the complete 2-electrode 3rd order Padé model was done using Maple software. The matrix expression is very lengthy, and symbolic computation of its rank is intractable. However, by plugging in typical parameter values, states, and inputs, the matrix can be computed numerically. The result is an observability matrix with full rank even for zero inputs (as long as the cell is not at steady state, i.e. $x_2, x_3, x_5, x_6 \neq 0$).