1   Parameter Estimation

We find the parameters $\mathbf{x}$ that solve the inverse problem $G(\mathbf{x}) = \mathbf{d}$, using a direct search method (*Nelder-Mead*).

$$\min \ \chi^2 = \min \ \sum_{i=1}^{n} \frac{(y_d(\mathbf{x})_i - y_G(\mathbf{x})_i)^2}{y_G(\mathbf{x})_i}$$

$$\mathbf{x} = \begin{cases} q \\ V_d \\ \sigma \end{cases} \quad \text{subject to constraints} \ \ g = \begin{cases} V_d & \pm & u_{exp} \\ \sigma & \pm & u_{exp} \\ y_0 & \pm & u_{exp} \\ t_0 & \pm & u_{exp} \end{cases}$$

where $y_G(\mathbf{x})$ is a numerical solution of the equation of motion

$$my'' = \frac{1}{2}\rho C_D A_d y'^2 + qE(y) + \frac{1}{4}\frac{Kq^2}{y^2} + \frac{1}{2}E(y)^2 \nabla \epsilon$$

In particular, it is impractical to directly measure the droplet net charge q, during a bounce experiement. Our workflow to identify these parameters is as follows:

1. Experimentally vary V_d, σ and capture droplet trajectories using a high-speed camera.

2. Digitize droplet trajectories by using automatic tracking of ellipse-fitted centroids on the thresholded video.

3. Slice droplet trajectories by their bounce minima, and apply a smoothing filter.

4. Maximize the goodness of fit between the experimental trajectories and solutions of the equation of motion, by varying the design vector (which encodes the unkown parameters in the equation of motion) using a direct search optimizer.

1.1 Inverse Problems

- We have a dynamical model describing the droplet electro-bounce; we wish to find typical values of the key parameter, droplet net charge q. In general it is not always possible nor practicle to measure model parameters experimentally. They may be hard, expensive, time consuming or perhaps even impossible to measure.

- A mathematical model designed to fit experiemental data so as to explicitly quantify physical parameters of interest.

- Values of model parameters are obtained using parameter estimation techniques aimed at providing a "best fit" to the data.

- Generally involves an interative process to minimize the average difference between the model and the data.

- Evaluating the quality of an inverse model involves a combination of established mathematical techniques as well as intuition and creative insight.

- A good inverse model has: good fit, model parameters are unique, and parameter values consistent with physical inuition or prior knowledge.

- The steps in the process are: 1. Select an appropriate mathematical model from theory. 2. Define a "figure of merit" function which measures agreement between the data and model for a given set of parameters. 3. Adjust model parameters to get a "best fit". 4. Evaluate "goodness of fit" to data, which tends never to be perfect due to experimental noise. 5. Estimate accuracy fo the best-fit parameter values (provide confidence intervals and determine uniqueness). 6. Determine whether a much better fit is possible (which is difficult of ruggest respose surfaces; F-test can be used for comparing models of different complexity).

- Maximum likelihood estimation: not "what is the probability that my set of model parameters is correct?" but rather "given my set of model

parameters, what is the probability that this data set occurred (what is the likelihood of the parameters given the data)?"

- Model parameter estimation is the process of indirect determination of unknown model parameters from measurements of experimental data.

- This is mathemetically challenging [ref]. There are practical challenges as well, as experimental data tends to be noisey and sparse.

- In the most familliar sense determining model parameters is done using derivative based methods, essentially by estimating derivatives by finite differences (Euler's method) and fitting paramters using linear regression. However there are approaches to fitting any arbitrary function.

- Suppose we have a model $G(\mathbf{m})$, with a vector of parameters $\mathbf{m}$, and set of perfect noiseless observations $\mathbf{d}$, we expect there to exist a relationship

$$G(\mathbf{m}) = \mathbf{d}$$

where the operator G might be an ODE.

- Suppose we have a model given by the ODE

$$\frac{dy}{dt} = f(t, \mathbf{y}; \mathbf{p}), \mathbf{y} \in \mathbb{R}^n, \mathbf{f} \in \mathbb{R}^n,$$

where $\mathbf{p} \in \mathbb{R}^m$ is the vector of parameters, and a collection of measurements of experiemental data

$$(t_1, \mathbf{y_1}), (t_2, \mathbf{y_2}), ..., (t_k, \mathbf{y_k}).$$

We wish to minimize an objection function which is the mean square error (or any goodness of fit figure of merit in general) between the model output

and experimental data:

$$\text{obj}(\mathbf{p}) = \sum_{j=1}^{k} |\mathbf{y}(t_j; \mathbf{p}) - \mathbf{y_j}|^2 \,,$$

where $|.|$ is the Euclidian norm. Therefore the parameter estimation problem is a variety of optimization problem.

- The process of fitting a function, defined by a collection of parameters, to a data set is called the discrete inverse, or parameter estimation problem (as opposed to the *forward problem* to find $\mathbf{d}$ given $\mathbf{m}$ and $G(\mathbf{m})$).

- By Bayes' Theorem

$$\text{prob}(X|D, I) = \frac{\text{prob}(D|X, I) \times \text{prob}(X|I)}{\text{prob}(D|I)}$$

where $\text{prob}(X|D, I)$ is the posterior probability density function that we want to estimate, $\text{prob}(D|X, I)$ is the likelihood function, $\text{prob}(X|I)$ is the prior probability density function that reflects our knowledge of the system, and $\text{prob}(D|I)$ is our evidence (the likelihood of the data based on our knowledge).

- The Maximum Likelyhood Estimate (MLE) for the the model prarameters $\mathbf{x_0}$, is given by the maximum of the posterior PDF, hwich is equivalent to the solution that produces the highest probability of the observed data. (The optimal parameter set is the one with the highest probability of observing the data and can be determined by maximizing the likelihood $\mathcal{L}$ of the data y with respect to the parameter set $\mathbf{x}$).

- The log-likelihood is given by

$$\mathcal{L} = \ln(\text{prob}(X|D, I)) = \text{const} - \frac{\chi^2}{2}$$

so the maximum of the posterior will occur when χ^2 is smallest, and the

corresponding optimal solution $\mathbf{x_0}$ is then called the least-squares estimate.

- We precondition the problem by minimizing $\ln(\chi^2)$, and scaling our constraints by their initial guesses.Scaling can help solve convergence problems and improve numerical stability. Here is goal is to make the problem qually senstive to steps in any direction. We use a naieve scaling (scaling variables such that their magnitudes ~ 1).

- Open questions: given the structure of a model is it possible to uniquely estimate it's unknown parameters? What experimental data are required to achieve unique parameter identification? These questions are called the problem of identifiability. Global identifiability is a challenging problem, and even in cases where is the problem is shown to be structurally identifiable, the problem may yet be practically unidentifiable. Our general approach to satisfying the uniqueness of our estimates is instead based on sensitivy analysis.

- the Hessian matrix must be negative definite for $\mathcal{L}$ to have a maximum at $\mathbf{x_0}$.

- We can use the condition number

$$\mathrm{cond}(A) = \frac{\max[\mathrm{eig}(A)]}{\min[\mathrm{eig}(A)]}$$

as a means to evaluate to stability of our problem. We find typical conditon numbers $\sim \mathcal{O}(10^{27})$ which indicate the problem is strongly ill-conditoned near the mimima $\mathbf{x_0}$.

- We're intested in the variance and co-variance as a means to determine the quality if the parameter estimate.

- The (i,j)-th element of the matrix $\sigma(\mathbf{x}, \mathbf{y})$ is equal to the covariance $\mathrm{cov}(X_i, Y_j)$ between the i-th scalar component of $\mathbf{x}$ and the j-th scalar component of $\mathbf{Y}$.

- Here the concept of error bars in linear correllation associated with a covariance matrix are not suitable. We might try to generalize the idea of confidence intervals to a multidimensional space, but usually it will be hard to describe the surface of the (smallest) hypervolume containing 90% of the probability in just a few numbers. The situation is worse if the probability density function has several maxima. However we notice that

$$\left[\sigma^2\right]_{ij} = -\left[\left(\nabla\nabla\mathcal{L}\right)^{-1}\right]_{ij} = 2\left[\nabla\nabla\left(\chi^2\right)\right]^{-1}_{ij} = -\left[H^{-1}\right]_{ij}$$

where H refers to the Hessian matrix and $\left[\sigma^2\right]_{ij}$ is the covariance matrix C. The likelhood function (and thus the postierior probability density function) are defined completely by the optimal solution $\mathbf{x}$ and the second derivative of $\mathcal{L}$ at the maximum, which corresponds to the covariance matrix C.

- The standard errors are the square roots of the diagonal of the covariance matrix. Our relative errors thus produced are extremely small ($> 1\%$). Verification of the results is crucial, yet it has not be done yet. [Ref] suggests verification by generation of Monte Carlo data sets. Does the estimation code recover the target parameters?

- We seek to maximize the likelyhood $\mathcal{L} = k - \frac{\chi^2}{2}$, which is equivalent to maximizing χ^2.

- Typically, each iteration evaluates the goal function only once or twice, which is why the Nelder-Mead algorithm is comparatively fast if goal function evaluation is the computational bottleneck (Singer and Mead, 2009).
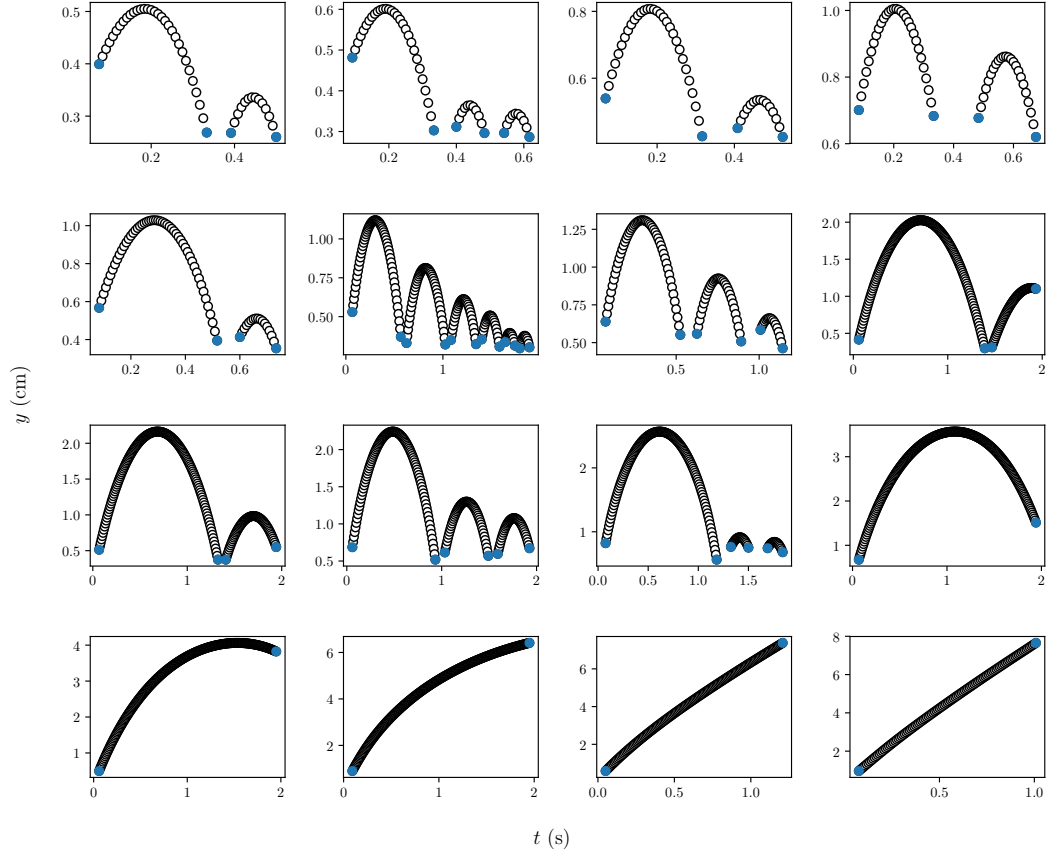
Figure 1: A simple EMA plot.

1.2 Optimization

Most generally the we state the constrained problem as

minimize: $F(\mathbf{x})$ objective function

subject to:

$g_j(\mathbf{x}) \leq 0$ inequality constraints
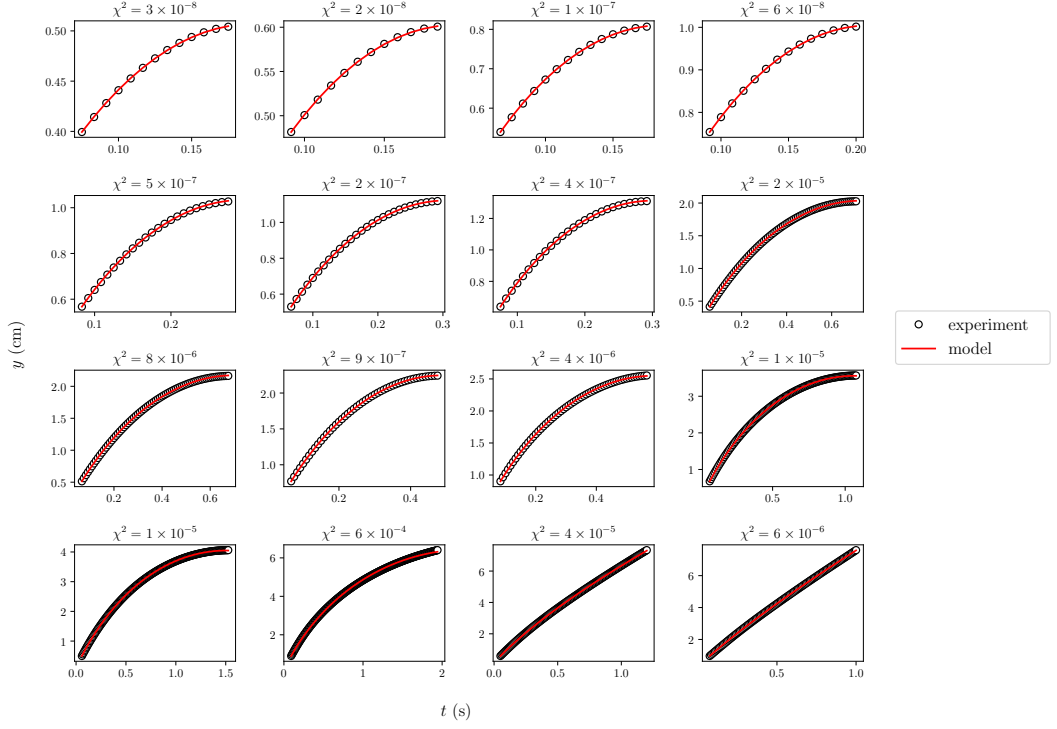
$h_k(\mathbf{x}) = 0$ equality constraints

Figure 2: A simple EMA plot.

$$\text{where } \mathbf{x} = \begin{cases} x_1 \\ x_2 \\ \vdots \\ x_n \end{cases} \quad \text{design variables}$$

Our specific optimization problem is non-convex, mixed discrete-continuously black-box (noisy), which is essentially the worst the worst case scenario. While in principle a gradient-based optimizer (such as ...) could be used by using finite-differences to obtain approximate gradients of the χ^2 objective funtion, in practice doing so is problematic becasue the signal-to-noise ratio of the objective function scales like $\mathcal{O}(f)$ for $\frac{df}{dt}$ and $\mathcal{O}(f^2)$ for $\frac{d^2 f}{ft^2}$ which will cause problems with our Hessians and Javobians respectively [Refs 5, 14, 49, 95.]. As a further practical matter, given the relatively expensive function-calls (which requires solving a stiff, non-linear ODE) gradient-free approaches tend to offer better
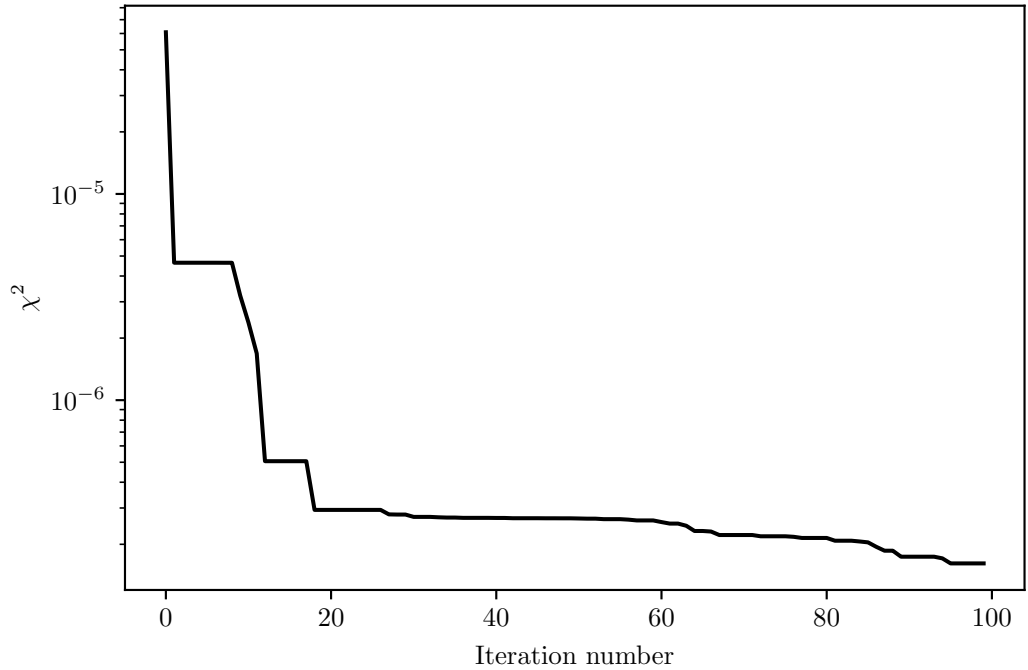
8

Figure 3: A simple EMA plot.

performance regardless.

Mathematical optimization is the problem of finding minima of a function f. In this context the function in question is referred to as a cost, or objective function. The field of mathematical optimization is as old as calculus itself, and the number of particular optimization techniques is correspondingly myriad; particular techniques lend themselves well to particular types of optimization problems. The minima of the objective function f is sought on a domain A specified by the constraints of the problem, this is usually called the feasible region. Minima of objective function $f : A \to \mathbb{R}^n$ on A are usually called feasible solutions. If the function f is convex the feasible solution is the global minimum, otherwise additional local minima exist. The scale of the optimization problem is set ultimately by dimensionality of the objective function. Functions may not always be smooth in the sense of having continuous derivatives, this is problematic in that optimization methods fundimentally rely on gradients of the

9

objective function. Problems with anisotropic objective functions in the sense that the gradient vector tends to differ significantly from the Newton direction $(-\mathbf{H}^{-1}f'(\mathbf{x})^T$, where $\mathbf{H}$ is the Hessian matrix) are considered ill-conditoned. In ill-conditioned problems gradient based deterministic search may be extremely slow. Numerical optimization may deal with black box functions (where we do not have an explicit mathematical expression of the function we are optimizing). Black box problems are challenging because we do not have access to analytic gradients of the objective function, and approximating them by finite-differences is slow and noisey. In general, noisey, black box, non-linear, non-quadratic, non-convex, constrained, ill-conditioned, high-dimensional objective functions are problematic to optimize. This is the practical essense of the parameter estimation problem. In particular there are few methods that work well with noisey, ill-conditioned problems.

We use a gradient-free, direct-search approach: Nelder-Mead [Nelder, 1965] implimented in `scipy.optimize` [Jones et al. 2001 –]. Nelder-Mead is robust to noise (relatively speaking), and is relatively thrifty with our extremely expensive function-calls. *Nelder-Mead*, sometimes called simplex-search or downhill-simplex, is a hueristic search method, with no guarantee of optimal solutions, but is well-established and widely used despite that. *Nelder-Mead* is based on the concept of a N-simplex, which generalizes a triangle into higher dimensions as a polytrope of $N+1$ vertices in N dimensions. It uses only-function calls and expands or contracts the simplex according to the function values at its vertices in a way visually reminiscent (in $\mathbb{R}^2$) of the oscillations of the jumping droplets themselves (in fact *Nelder-Mead* is sometimes also called the "ameoba method"). Very little is known about the convergence properties of the *Nelder-Mead* algorithem in its classical form for non smooth objective functions ([Price and Coope, 2003]), except that in general it doesnt satisfy the properties required for convergence by other direct search algorithems: that the simplex remains uniformly non-degenerate, and that some form of "sufficient" descent conditon for function values at the vertices is required at each iteration

We do not have *a priori* information regarding existence of true globality for our objective function. Perhaps more importantly, given the degrees-of-freedom in the parameter space there may be a multiplicity of local extrema of the χ^2 hyper-response surface. *Nelder-Mead* is not a global optimizer, though there are varients which use sequential local searches with probablistic restarts to achieve globality. However global optimization usually comes at a tremendous computational cost.

That we are capable of fitting any arbitrary model to a dataset given sufficient degrees of freedom in our parameters is admittedly a disconcerting issue, and is manifested in the locality of the extrema of the χ^2 response surface. However, some of the inverse model parameters are constrained by our experimental observations of them and their associated measurement uncertainties. This, we hope, makes the spectre of an overfitted model less frightening, but does convert our uncontrained optimization problem to an constrained one which raises special difficulties of its own.

The *Nelder-Mead* direct search method cannot be used with explicity constrained problems. However, there are various implicit approaches to (approximately) solving general constrained problems using uncontrained algorithems. Generally, this is achieved by domain transformations or the use of penalty functions. By the addition of a penalty function which depends in some way on the values of the constraints to the objective function, we minimize a pseudo-objective function where the infeasibility of the constraints is minimized simultaneously to the objective function.

There are various penalty function schemes. We use an Exterior Penalty Function as a simple way of converting converting the constrained problem into an unconstrained one. These are are especially useful in cases where the constraints are not "hard" in the sense that they need to be satisfied precisely. General penalty functions, which are sequential unconstrained minimization techniques, reformulate the general constrained problem as the pseudo-objective function given by

$$\phi(\mathbf{x}, r_p) = F(\mathbf{x}) + r_p P(\mathbf{x})$$

where $P(\mathbf{x})$ the penalty function, is given by

$$P(\mathbf{x}) = \sum_{j=1}^{m} \{\max[0, g_j(\mathbf{x})]\}^2 + \sum_{k=1}^{l} [h_k(\mathbf{x})]^2. \tag{1}$$

We see from Equation 1 that there is no penalty if the constraints $g_j(\mathbf{x})$, and $h_k(\mathbf{x})$ are satisfied.

The Exterior Penalty Function specifically (and all Penalty Function approaches in general) do have several drawbacks. Namely these include the possibility of the objective function being undefined outside of the set of feasible solutions. Additionaly, by naively "encouraging" feasibility of the solution using large values of the penalty parameter, r_p, we will tend to ill-condition the unconstrained formulation of the problem (thoug in our implimentation the preconditoning tends to make the psuedo-objectibve function less and less sentive to the constraints as the likelihood is approaches a maximum). [notes on choice of optimizer parameters, initial guesses, convergence, error estimates]. Using penalty function methods tend to exascerbate the sensitivity of *Nelder-Mead* solutions to the choice of intial guess vector [ref].

Despite their wide use in parameter estimation, there is very little known about convergence properties of *Nelder-Mead.* The iterative procedure may thus converge very slowly near the optimum, unless the Hessian is "well-conditioned", i.e., the condition number is near one. However, if parameter estimates are highly correlated among themselves, the Fisher information matrix is near-singular and its inversion is either impossible or involves significant numerical error.

1.3 Smoothing

In trying to recover the kinematic variables from the droplet trajectories we encounter a difficulty in the noise of the position data:

Error sources include misalignment of the camera, perspective due to objects (subject or reference scale) being out of the photographic plane, precision limits in digitization. Some of these errors are systematic in origin and introduce consistent biases into the data (e.g. coherent spectral sources, rather than truly stochastic noise). These systematic sources of error include inaccurate scales among others. Data smoothing tends not to help with the systematic errors in that they are usually of lower frequency than the signal (and here we are trying specifically to filter high frequency noise). Random errors, by contrast, are assumed to have a Gaussian distribution (by the central limit theorem), and are independent of the signal (which inherently results from a deterministic process).

We experimented with a variety of filters implimented in the `scipy.signal` *SciPy*[ref] module using a represenative set of trajectory data; these methods include 1D Gaussian convolution, Wiener, Butterworth, and Savitsky-Golay filters. Qualitatively comparing these smoothing methods (by handtuning filter orders and window sizes) we find that we loose too many data points in the smoothing process, large amplitudes are overly smoothed by repeated filtering passes, or there are significant end effects for most of these methods. A comparison of these smoothing approaches on a representative trajectory data set are shown in Figure **??**. The power spectra for the same data are compared for these methods in Figure **??**.

Qualitively speaking, the 4th-order Savitsky-Golay, and Butterworth filters both produce the smooth derivatives we desire for for the optimization algorithem; but the small window-size needed for Butterworth filter tends to also prduce a noticible end effect. The Savitsky-Golay filter essentially uses a moving-window based on local least-squares polynomial approximations. It was shown that fitting a polynomial to a set of input samples and then evaluating
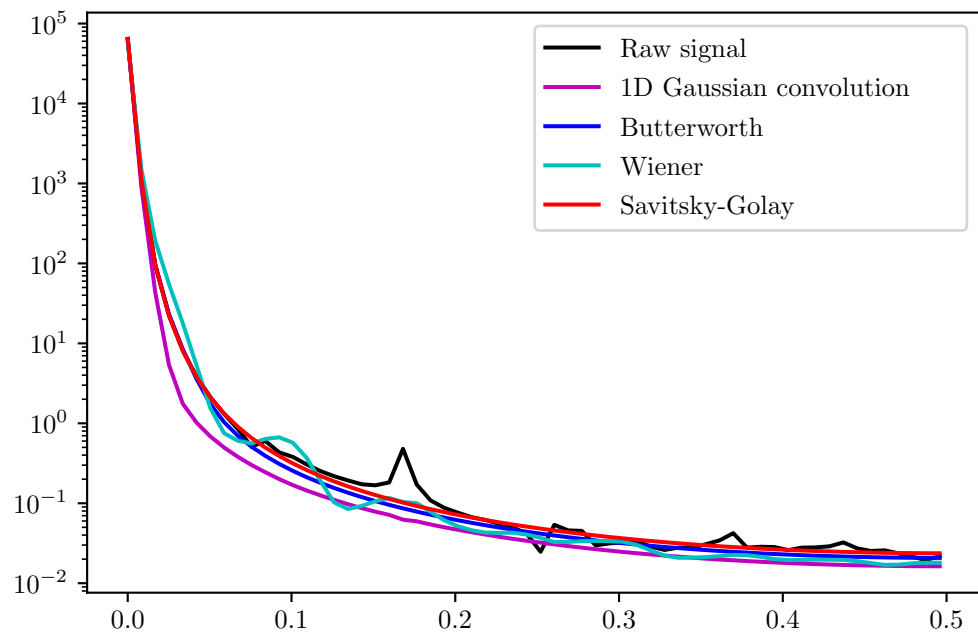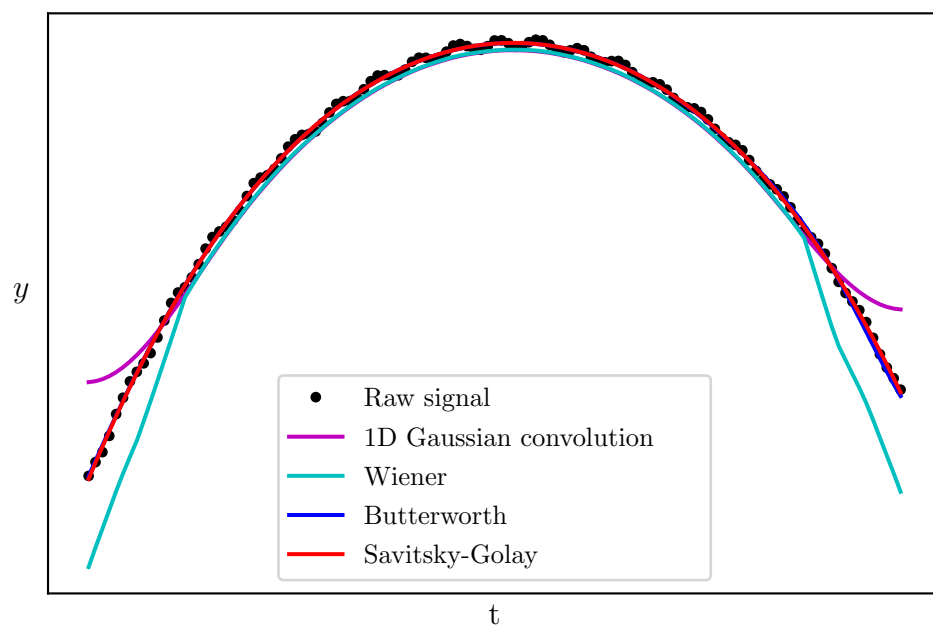
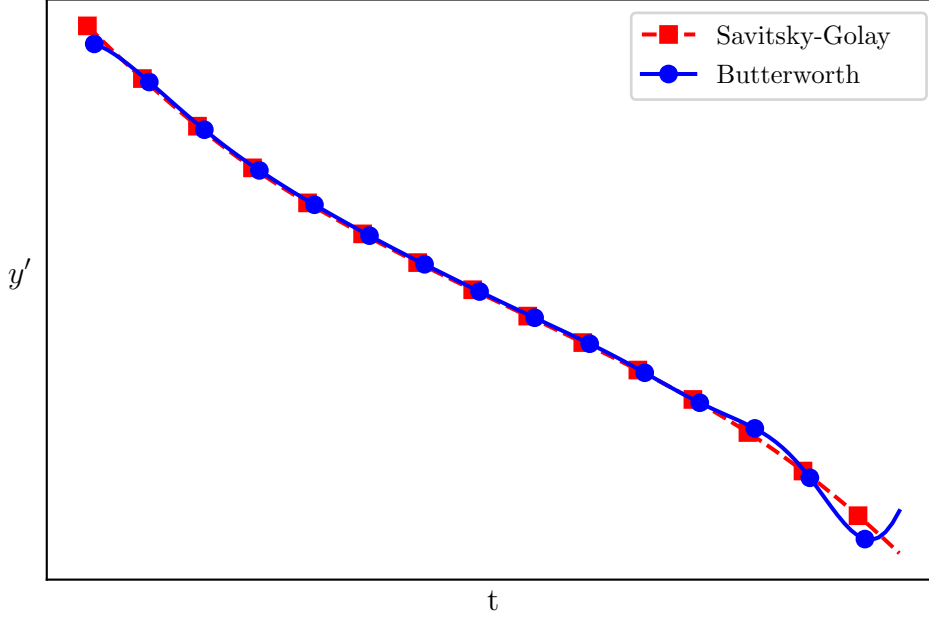Figure 4: A simple EMA plot.

Figure 5: A simple EMA plot.

Figure 6: A simple EMA plot.

the resulting polynomial at a single point within the approximation interval is equivalent to discrete convolution with a fixed impulse response [Savitsky, 1964]. One property of this kind of low-pass filter is their tendency to respect waveform aplitudes, and so they are attractive in applications with having noisey signals with shaply pointed waveforms such as ultrsound or synthetic aperature radar [Schafer, 2011], however these filters tend to suffer from end-effects. Because Savitsky-Golay is a Finite Impulse Response (FIR) filter it requires datapoints to be equally spaced; to accomedate this we interpolate points between the small gaps which sometime occur in the tracking results from image analysis [ref]. [Notes on window length, polynomial order].

1.4 Numerical Solution to the Equation of Motion

The equation of motion behaves stiffly due to the large disparity in Coulombic, image charge, and dielectrophoretic lenghtscales. We integrate it numerically using the **odeint** *Scipy* module. This is a shake-and-bake Python wrapper for the

vernerable 1982 *netlib ODEPACK* library double-precision `lsoda` (Livermore Solver for Ordinary Differential equations with Automatic method switching for stiff and nonstiff problems) integrator [ref]. The function switches between Adams (nonstiff) and Backwards Differentiation Formulas (BDF, stiff) according to the dyanmic value of a set of stiffness eigenvalues.