Regression

In this exercise you will be considering three different regression techniques ($\mathbf{x} \in \mathbb{R}^N$ and $y \in \mathbb{R}$), namely:

1. **Regular Least Squares (RLS):**
   - Regressor: $y = w^T\mathbf{x} + b$
   - Optimisation: $w = (XX^T)^{-1}Xy$
2. **Weighted Least Squares (WLS):**
   - Regressor: $y = w^T\mathbf{x} + b$
   - Optimisation: $w = (ZZ^T)^{-1}Zv$ where $Z = XB^{1/2}$ and $v = B^{1/2}y$
3. **Locally Weighted Regression (LWR):**
   - Regressor: $y = \left(\sum_{i=1}^{M} \beta_i(\mathbf{x})y^i\right) / \left(\sum_{i=1}^{M} \beta_i(\mathbf{x})\right)$

     The beta is a kernel density function centred on a point i: $\beta_i(\mathbf{x}) = \exp(-\frac{1}{2}\|\mathbf{x}^i - \mathbf{x}\|^{\frac{1}{2}})$
   - Optimisation: no-optimisation, data driven.

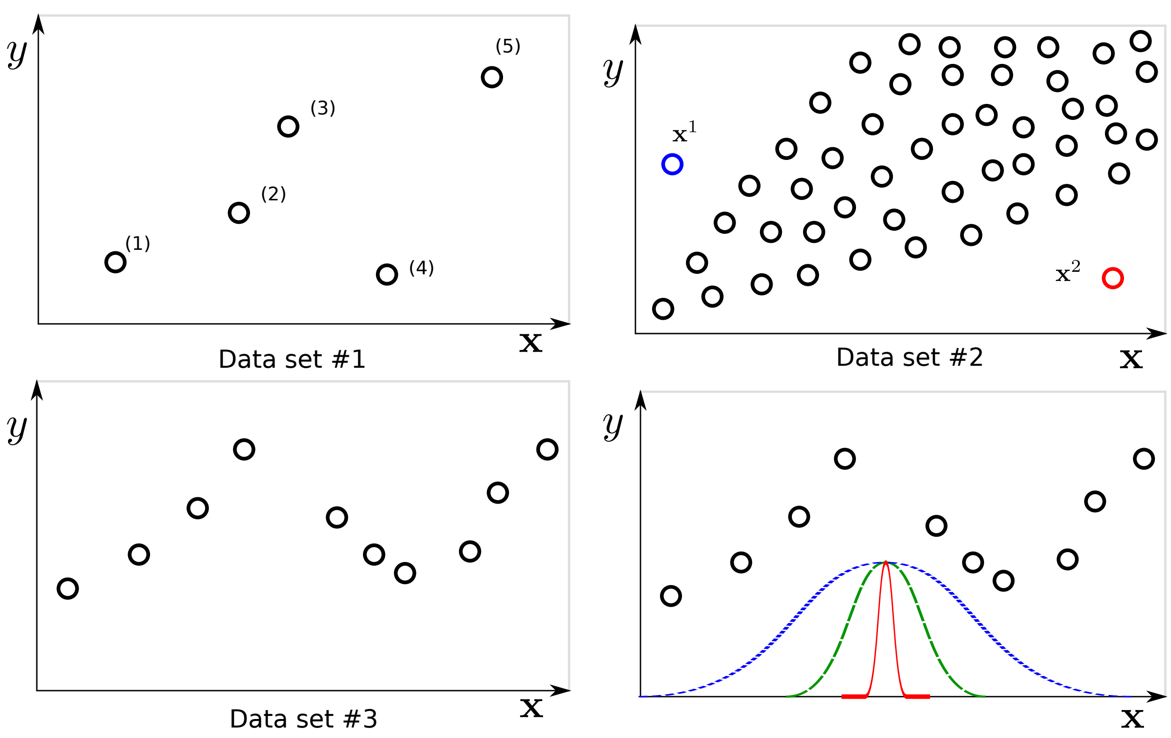


Figure 1: Three datasets, the black circles depict the data points. $\mathbf{x}$ is the input and y is the output and we wish to estimate $y = f(\mathbf{x})$.

A) In Figure 1, three different datasets are given

1. Draw the solution that RLS would give you for datasets 1 to 3 (do not consider the colored points in dataset 2).
2. Given the set of weights $\boldsymbol{\beta} = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, \frac{1}{4}]$, apply WLS to dataset 1 and draw the resulting regression function.
3. What solution WLS would give for dataset 2, considering that the blue (point $\mathbf{x}^1$) and the red data point (point $\mathbf{x}^2$) are weighted with $\beta(\mathbf{x}) = \frac{1}{\mathbf{x}}$.
4. Draw the solutions of LWR for dataset 3 with each of the given kernels (see Figure 1, *Bottom right*).

B) Your lab (Lab 1) is studying a rare type of particles. Using particles with different sizes your lab took measurements of their speed. You wanted more data so you asked a cooperating lab (Lab 2) to share their measurements with you (figure 2). Lab 1 was using a measuring instrument with the Gaussian error $e_1 \sim \mathcal{N}(0, 10)$, while the error of Lab 2 measurements was $e_2 \sim \mathcal{N}(0, 20)$. You want to find out what's the linear relation between the speed of a particle and its size. Which regression method should you use, and how would you use it?

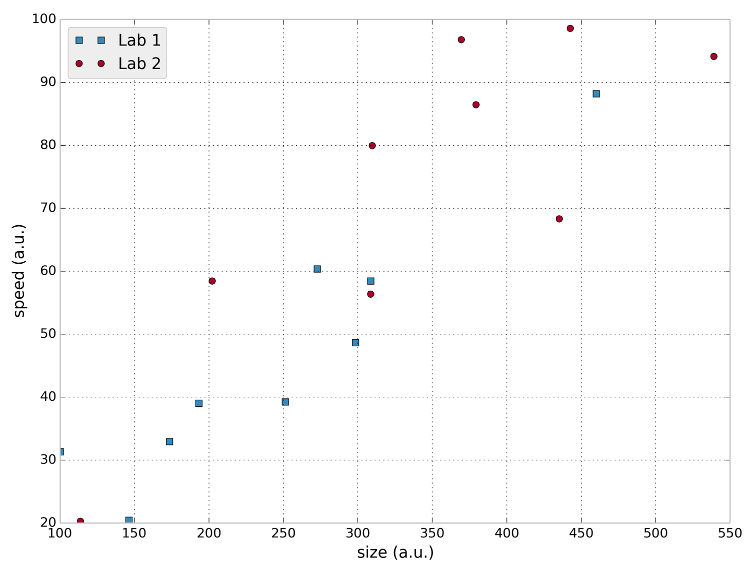


Figure 2: The datapoints collected from the two labs.

Solutions

A) For notation purposes, $e_i = \hat{y}^i - y^i$ is the error between the true value y (training data) and the estimated value $\hat{y}$, from the regressor function.

1. All the resulting regression lines minimize the sum of square errors (see Figure 3). This is the sum of the red distances between the predicted and actual values of y.

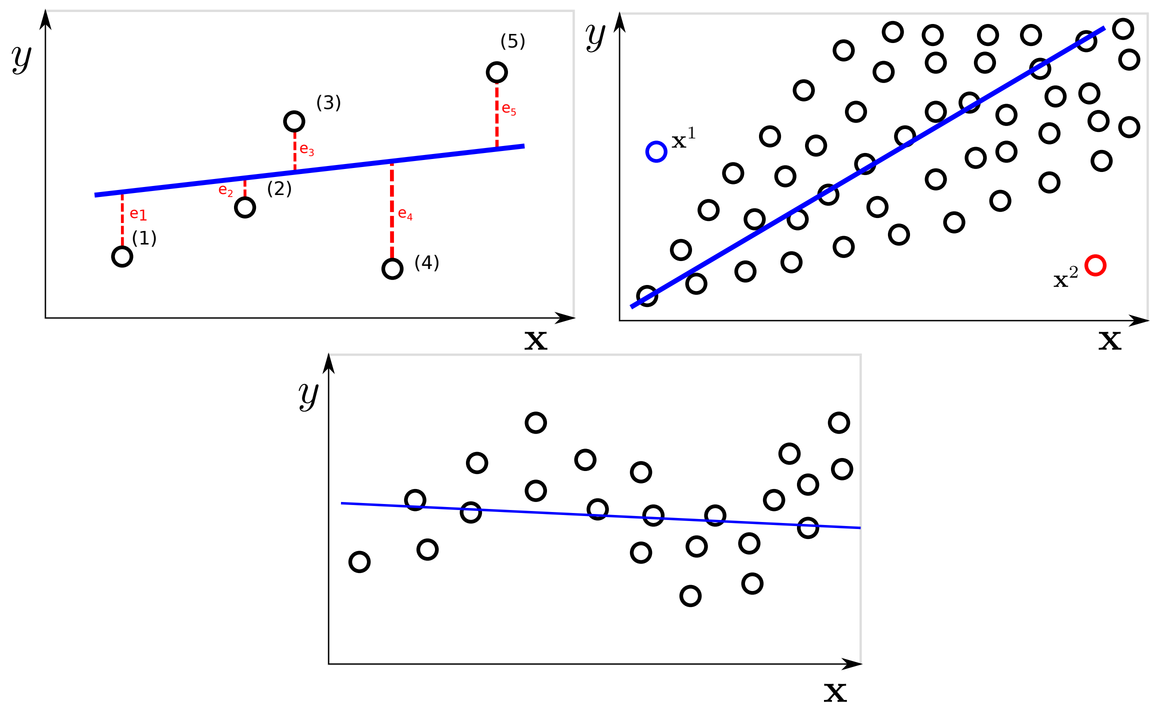


Figure 3: Regression functions obtained using RLS regression on the first three datasets.

2. Because the weight for the 4th data point is zero, it plays no role in the optimisation of the weights, W. As a result the line ignores this point (see Figure 4).

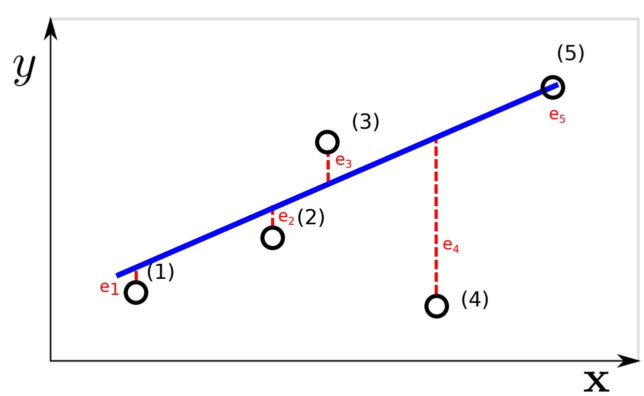
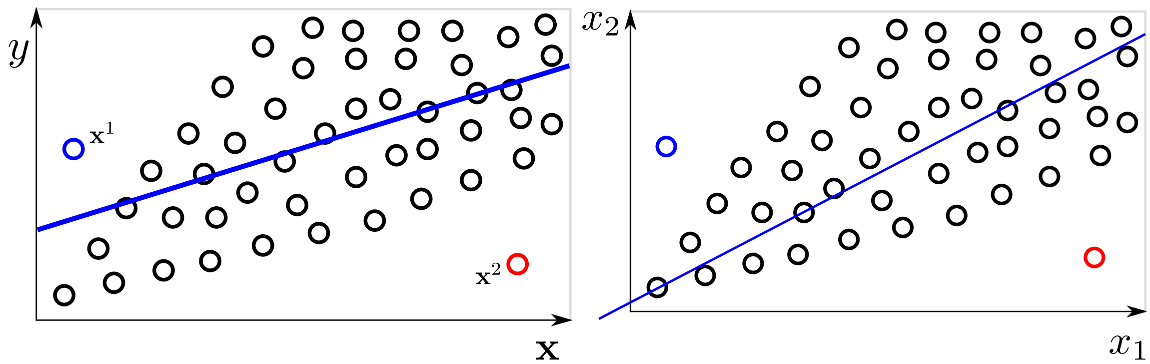


Figure 4: Regression function obtained using WLS regression on dataset 1 with $\boldsymbol{\beta} = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, \frac{1}{4}]$

3. The functional form of the beta implies that as points are located further away from the origin they will impact less the final regression line (see Figure 5a). The datapoint $\mathbf{x}^2$ being far away will thus not influence the line. However point $\mathbf{x}^1$ will. Using identical weights for all points would in turn yield the regression function shown in Figure 5b.



(a) (b)

Figure 5: Regression function obtained using WLS regression on dataset 2 (a) with $\beta(\mathbf{x}) = \frac{1}{\mathbf{x}}$ for the red and blue points, (b) with equal weightings.

4. The regression functions obtained using LWR for dataset 3 with each of the given kernels are displayed are displayed in Figure 6.

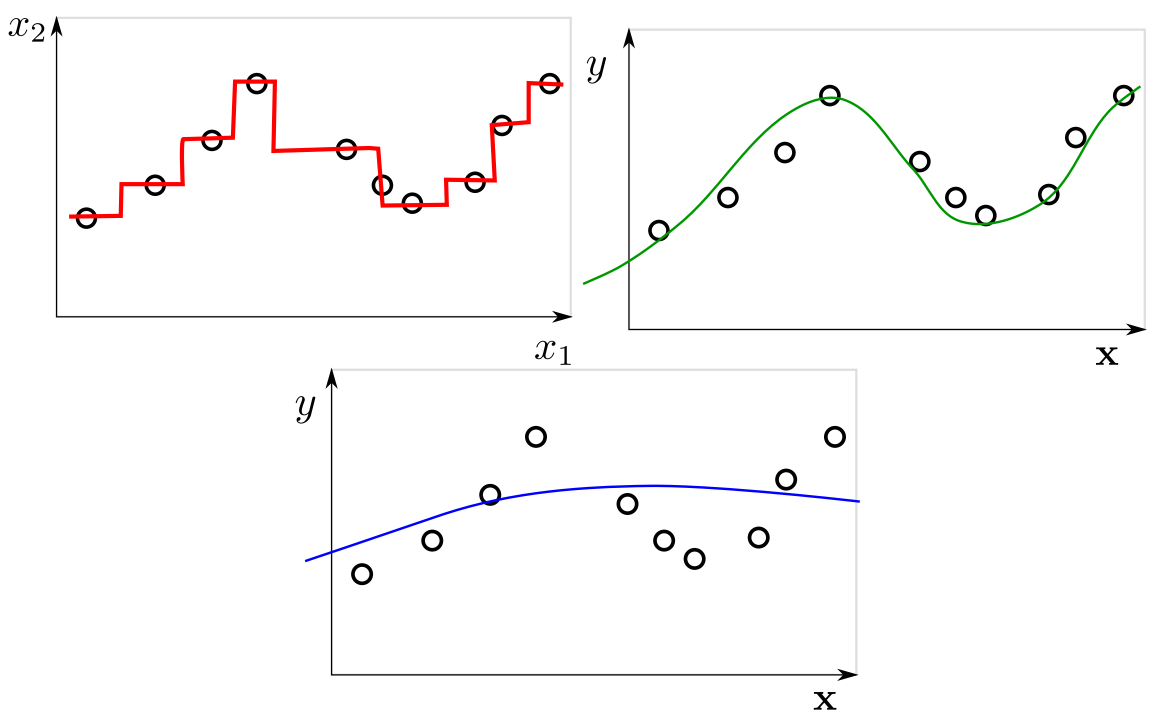


Figure 6: Regression functions obtained using LWR for dataset 3 with each kernel given in Figure 1 (bottom right)

As an another example, we give results obtained in Matlab on another dataset for 3 different kernel widths.

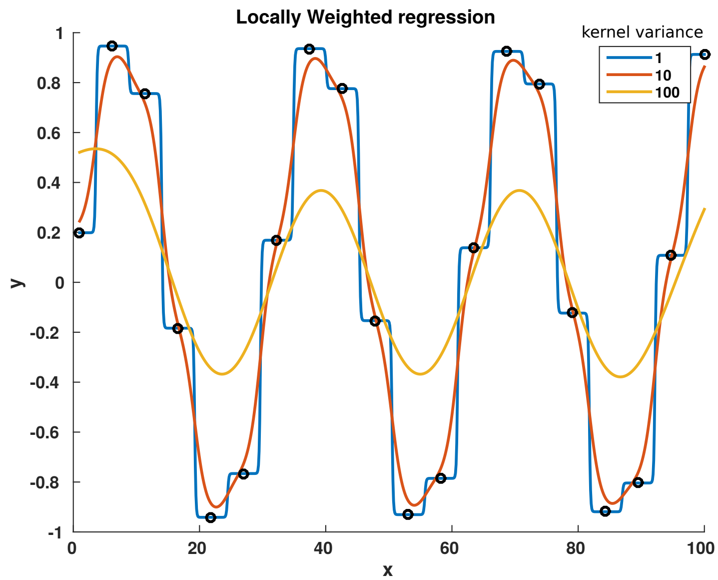


Figure 7

B) The measurements from Lab 1 have a smaller error than the measurements from Lab 2. For this reason we should give more credibility (weight) to Lab 1 measurements. Therefore, we should use the Weighted Least Squares to solve this problem (see Figure 8).
Note: It has been proven that the best linear unbiased estimation is computed when the diagonal entries of the weight matrix, B, are equal to the inverse of the variance of the measurements, i.e.
$$\beta_i = \begin{cases} \frac{1}{\sigma_1^2}, & \text{if datapoint } i \text{ from Lab 1} \\ \frac{1}{\sigma_2^2}, & \text{if datapoint } i \text{ from Lab 2} \end{cases}, \forall i = 1..M.$$

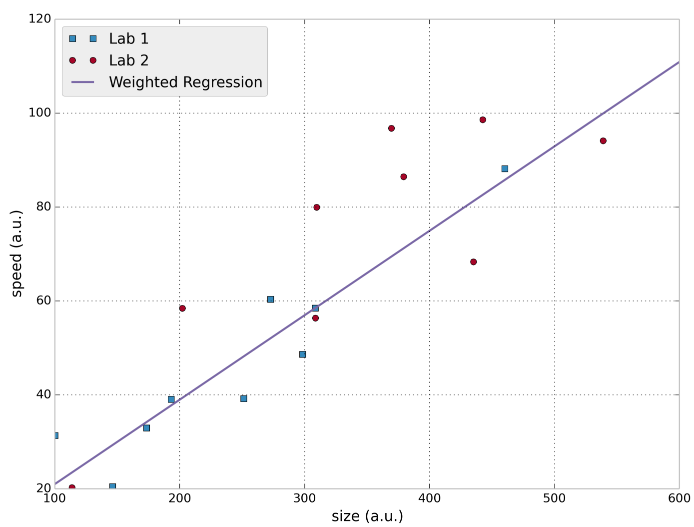


Figure 8: The datapoints collected from the two labs and the obtained regression line.

Least Squares

In the lecture you have covered linear function estimators of the following form:

$$y^i = f(\mathbf{x}^i; \mathbf{w}, b) = \mathbf{w}^{\mathrm{T}}\mathbf{x}^i + b \tag{1}$$

where $\mathbf{w} \in \mathbb{R}^N$ and $\mathbf{x} \in \mathbb{R}^N$ are $(N \times 1)$ column vectors, b is the scalar intercept and y is the predictor.

Given you have a set of M data points, $X = [\mathbf{x}^1, \dots, \mathbf{x}^i, \dots, \mathbf{x}^M]$, and associated predictors, $\mathbf{y} = [y^1, \dots, y^i, \dots, y^M]$. Consider the Sum of Squared Error (SSE) as your loss function and derive the optimal choice of parameters of the linear regressor for the **bivariate** case:

$$y^i = wx^i + b \tag{2}$$

$$SSE = \sum_{i=1}^{M} \left(y^i - f(x^i)\right)^2 = \sum_{i=1}^{M} e_i^2 \tag{3}$$

where e_i is the error between the target and predicted value.

Solutions

The solution for $\mathbf{w}$ is obtained by minimizing the SSE by taking the derivatives with respect to there parameters in question.

$$J(w,b) = \sum_{i=1}^{M} \left(y^i - wx^i - b\right)^2$$

$$\begin{aligned}
\frac{\partial J(w,b)}{\partial b} &= -2\sum_{i=1}^{M} \left(y^i - wx^i - b\right) = 0 \\
&= -2\left(\sum_{i=1}^{M} y^i - wx^i\right) + 2Mb \\
\implies b &= \frac{1}{M}\left(\sum_{i=1}^{M} y^i - wx^i\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial J(w,b)}{\partial w} &= -2\sum_{i=1}^{M} \left(y^i - wx^i - b\right) x^i = 0 \\
&= -2\sum_{i=1}^{M} y^i x^i + 2w\sum_{i=1}^{M} (x^i)^2 + 2b\sum_{i=1}^{M} x^i \\
\implies w &= \left(\sum_{i=1}^{M} x^i(x^i - \overline{x})\right)^{-1}\left(\sum_{i=1}^{M} (y^i - \overline{y})x^i\right)
\end{aligned}$$

where $\overline{x} = \frac{1}{M}\sum_{j=1}^{M} x^j$ and $\overline{y} = \frac{1}{M}\sum_{j=1}^{M} y^j$.

Control of Robotic Manipulator (to be done at home)

Consider the 3 degree of freedom, $\mathbf{q} = \{q_1, q_2, q_3\}$, robotic arm in Fig. 9. The vector $\mathbf{q}$ denotes the current joints' position while $\mathbf{x} \in \mathbb{R}^2$ is the location in the 2D space of the tip of the robotic arm, also known as end-effector. The position of the end-effector is connected to the joints' position

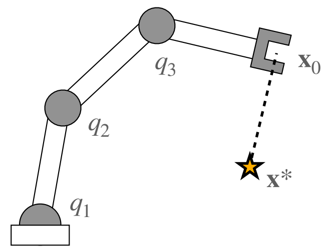


Figure 9: 3 degree of freedom robotic manipulator.

through the forward kinematics equation $\mathbf{x} = \phi(\mathbf{q})$.

A) We are interested in generating a joints velocity vector, $\dot{\mathbf{q}}$, that would move the end-effector of the robot from the current location $\mathbf{x}_0$ towards the goal location $\mathbf{x}^*$. Show that the optimal $\dot{\mathbf{q}}$ is the solution of a least-square linear (unweighted) regression of the form $\mathbf{w} = (XX^T)^{-1}X^T\mathbf{y}$ (Hint: the derivative of the forward kinematics with respect to the joints' position is $J(\mathbf{q}) = \frac{\partial\phi(\mathbf{q})}{\partial\mathbf{q}}$, also know as Jacobian).

B) We would like to move the first joint of the robot without changing the current location of the end-effector. Is there any other joints velocity vector $\dot{\mathbf{q}}$, solution of the linear regression problem derived in the previous step, that would achieve this?

Solutions

A) In order to move the end-effector tip towards $\mathbf{x}^*$ we want to generate a velocity of the end-effector proportional to the distance between the end-effector goal and current location $\dot{\mathbf{x}} \propto \mathbf{x}^* - \mathbf{x}_0$. Deriving the forward kinematics in time, $\frac{d}{dt}$, on both side of the equation we obtain $\dot{\mathbf{x}} = J(\mathbf{q})\dot{\mathbf{q}}$. This is a linear regression problem where $J(\mathbf{q})$ plays the role of the dataset, $\dot{\mathbf{x}}$ is the ground truth and $\dot{\mathbf{q}}$ are the unknown (weights) that we want to determine. The least mean square solution of such a problem is $\dot{\mathbf{q}} = (J(\mathbf{q})^T J(\mathbf{q}))^{-1} J(\mathbf{q})^T (\mathbf{x}^* - \mathbf{x}_0)$.

B) We are now interested in keeping the position of the end-effector fixed, therefore $\dot{\mathbf{x}} = 0$. Nevertheless we would like to move the first joint of the robot while preserving, as much as possible, the current position of the second and third joint. Observe that $J(\mathbf{q}) \in \mathbb{R}^{2\times 3}$. This is equivalent to consider a regression problem where we have 2 data points in 3 dimensions. This yields an overdetermined problem where the unknown variables is fewer than the constraint equations. Therefore $\mathrm{rank}(J(\mathbf{q})) = 2$. This leads to infinite solutions of the type $\dot{\mathbf{q}} = (J(\mathbf{q})^T J(\mathbf{q}))^{-1} J(\mathbf{q})^T \dot{\mathbf{x}} + \dot{\mathbf{q}}_n$ with $J(\mathbf{q})\dot{\mathbf{q}}_n = 0$. This means that $\dot{\mathbf{q}}_n$ belongs to the null space of J. We would like to move the first joint, therefore $\dot{\mathbf{q}}_d = [\alpha, 0, 0]^T$, with α some constant. In order to project $\dot{\mathbf{q}}_d$ in the null-space of $J(\mathbf{q})$ we need to determine the projector operator P such that $J(\mathbf{q})P\dot{\mathbf{q}}_d = 0$. It is possible to verify that $P = I - (J(\mathbf{q})^T J(\mathbf{q}))^{-1} J(\mathbf{q})^T J(\mathbf{q})$ is the projector operator into the null-space of $J(\mathbf{q})$. Therefore our solution is $\dot{\mathbf{q}} = P\dot{\mathbf{q}}_d$ since the first term cancels out given that $\dot{\mathbf{x}} = 0$.