l_1 Trend Filtering

name: Jaehyun Lim

Student ID: 2017311490

Introduction

Trend Filtering: Estimate a trend component $x_t, t = 1, \ldots, n$, (slowly varying) from y_t where $y_t = x_t + z_t$. z_t is a random component (rapidly varying compared to x_t)

Application: macroeconomics, geophysics, financial time series analysis, etc…

Previous Methods: Hodrick-Prescott (H-P) filtering, moving average filtering, bandpass filtering, smoothing splines…

Hodrick-Prescott filtering

the trend estimate with regularization:

$$\frac{1}{2}\sum_{t=1}^{n}(y_t - x_t)^2 + \lambda\sum_{t=2}^{n-1}(x_{t-1} - 2x_t + x_{t+1})^2$$

Regularizing the square sum of the second-order difference.

The objective function can be written in matrix form as

$$\frac{1}{2}\|y - x\|_2^2 + \lambda\|Dx\|_2^2$$

Where $x \in \mathbb{R}^n, y \in \mathbb{R}^n, D \in \mathbb{R}^{(n-2)\times n}$ is the second-order difference matrix

$$D = \begin{bmatrix} 1 & -2 & 1 & & & & \\ & & & \ddots & & & \\ & & & & 1 & -2 & 1 \end{bmatrix}$$

The analytic solution of H-P trend estimate is

$$x^{hp} = \left(I - 2\lambda D^T D\right)^{-1} y$$

l_1 trend filtering

Variation of H-P filtering which regularize the sum of absolute value of the second-order difference of the time series values.

$$\frac{1}{2}\sum_{t=1}^{n}(y_t - x_t)^2 + \lambda\sum_{t=2}^{n-1}|x_{t-1} - 2x_t + x_{t+1}|$$

Which can be written in matrix form as

$$\frac{1}{2}\|y - x\|_2^2 + \lambda\|Dx\|_1$$

Properties of l_1 trend filtering

$O(n)$ arithmetic operations

$x^{lt} \to y$ as $\lambda \to 0$

x^{lt} is not a linear function of the original data y (x^{hp} is a linear function of y)

$x^{lt} = x^{ba}$ for$\lambda = \lambda_{max} \geq \left\| \left(DD^T \right)^{-1} Dy \right\|_\infty$

$x^{lt} = \frac{\lambda_i - \lambda}{\lambda_i - \lambda_{i+1}} x^{(i+1)} + \frac{\lambda - \lambda_{i+1}}{\lambda_i - \lambda_{i+1}} x^{(i)}, \lambda_{i+1} \leq \lambda \leq \lambda_i, i = 1, \dots, k-1$, where $x^{(i)}$ is x^{lt} with $\lambda = \lambda_i$, $\lambda_1 = 0, \lambda_k = \lambda_{max}$

Let $\tilde{x}^{lt}$ denote the l_1 trend estimate for $(y_1, \dots, y_{n+1})$. There is an interval $[l, u]$ with $l < u$, for which $\tilde{x}^{lt} = \left(x^{lt}, 2x_n^{lt} - x_{n-1}^{lt} \right)$, provided $y_{n+1} \in [u, l]$

Properties of l_1 trend filtering

l_1 trend filtering problem is equivalent to the l_1 regularized least squares problem

$$\text{minimize} \frac{1}{2} \|A\theta - y\|_2^2 + \lambda \sum_{i=3}^{n} |\theta_i|$$

Where $\theta = (\theta_1, \dots, \theta_n) \in \mathbb{R}^n$ is the variable and A is the lower triangular matrix

$$A = \begin{bmatrix} 1 & & & & & \\ 1 & 1 & & & & \\ 1 & 2 & 1 & & & \\ 1 & 3 & 2 & \ddots & & \\ \vdots & \vdots & \vdots & \ddots & 1 & \\ 1 & n-1 & n-2 & \dots & 2 & 1 \end{bmatrix} \in \mathbb{R}^{n \times n}$$

From a standard result in l_1-regularized least squares, the solution θ is a piecewise-linear function of the regularization parameter λ

Optimality condition

First-order optimality condition based on subdifferential calculus

$$\exists \nu \in \mathbb{R}^{n-2} \text{ s.t. } y - x = D^T \nu,\ \ \nu_t \in \begin{cases} \{+\lambda\},\ (Dx)_t > 0, \\ \{-\lambda\},\ (Dx)_t < 0, \\ [-\lambda, \lambda],\ (Dx)_t = 0, \end{cases} \quad t = 1, \ldots, n-2.$$

can be written as

$$\left(\left(DD^T \right)^{-1} D(y - x) \right)_t \in \begin{cases} \{+\lambda\}, & (Dx)_t > 0, \\ \{-\lambda\}, & (Dx)_t < 0, \\ [-\lambda, \lambda], & (Dx)_t = 0, \end{cases} \quad t = 1, \ldots, n-2.$$

Dual problem

Reformulation with new variable $z \in \mathbb{R}^{n-2}$ and new equality constraint $z = Dx$

$$\text{minimize} \frac{1}{2}\|y-x\|_2^2 + \lambda\|z\|_1$$
$$\text{subject to } z = Dx$$

The Lagrangian is

$$L(x,z,\nu) = \frac{1}{2}\|y-x\|_2^2 + \lambda\|z\|_1 + \nu^T(Dx - z)$$

The dual function is

$$\inf_{x,z} L(x,z,\nu) = \begin{cases} -\frac{1}{2}\nu^T DD^T\nu + y^T D^T \nu, & -\lambda\mathbf{1} \le \nu \le \lambda\mathbf{1}, \\ -\infty, & \text{otherwise} \end{cases}$$

The dual problem is

$$\text{minimize } g(\nu) = \frac{1}{2}\nu^T DD^T\nu - y^T D^T\nu$$
$$\text{subject to } -\lambda\mathbf{1} \le \nu \le \lambda\mathbf{1}$$

from the solution ν^{lt} of the dual problem, $x^{lt} = y - D^T\nu^{lt}$.

Dual and central residual

$$r_t(\nu, \mu_1, \mu_2) = \begin{bmatrix} r_{dual} \\ r_{cent} \end{bmatrix} =$$

$$\begin{bmatrix} \nabla g(\nu) + D(\nu - \lambda\mathbf{1})^T \mu_1 - D(\nu + \lambda\mathbf{1})^T \mu_2 \\ -\mu_1(\nu - \lambda\mathbf{1}) + \mu_2(\nu + \lambda\mathbf{1}) - \left(\frac{1}{t}\right)\mathbf{1} \end{bmatrix}$$

is the residual, where $\mu_1, \mu_2 \in \mathbb{R}^{n-2}$ are dual variable for the inequality constraint. As $t \to \infty, r_t(\nu, \mu_1, \mu_2) = 0$ reduces to the KKT-condition.

Primal-dual interior point method

The Newton step is characterized by

$$
\begin{aligned}
&r_t(\nu + \Delta\nu, \mu_1 + \Delta\mu_1, \mu_2 + \Delta\mu_2) \\
&\approx r_t(\nu, \mu_1, \mu_2) + Dr_t(\nu, \mu_1, \mu_2)(\Delta\nu, \Delta\mu_1 \Delta\mu_2) = 0
\end{aligned}
$$

Where Dr_t is the derivative of r_t. This can be written as

$$
\begin{bmatrix} DD^T & I & -I \\ I & J_1 & 0 \\ -I & 0 & J_2 \end{bmatrix} \begin{bmatrix} \Delta\nu \\ \Delta\mu_1 \\ \Delta\mu_2 \end{bmatrix} = \begin{bmatrix} DD^T z - Dy + \mu_1 - \mu_2 \\ f_1 + \left(\frac{1}{t}\right) \mathrm{diag}(\mu_1)^{-1} \mathbf{1} \\ f_2 + \left(\frac{1}{t}\right) \mathrm{diag}(\mu_2)^{-1} \mathbf{1} \end{bmatrix},
$$

where

$$
\begin{gathered}
f_1 = \nu - \lambda \mathbf{1} \in \mathbb{R}^{n-2} \\
f_2 = -\nu - \lambda \mathbf{1} \in \mathbb{R}^{n-2} \\
J_i = \mathrm{diag}(\mu_i)^{-1} \mathrm{diag}(f_i) \in \mathbb{R}^{(n-2)\times(n-2)}
\end{gathered}
$$

Eliminating $(\Delta\mu_1, \Delta\mu_2)$,

By eliminating $(\Delta\mu_1, \Delta\mu_2)$, the reduced system is:

$$\left(DD^T - J_1^{-1} + J_2^{-1}\right)\Delta\nu$$
$$= -\left(DD^T z - Dy - \left(\frac{1}{t}\right)\mathrm{diag}(f_1)^{-1}\mathbf{1} + \left(\frac{1}{t}\right)\mathrm{diag}(f_2)^{-1}\mathbf{1}\right)$$

$\Delta\mu_1$ and $\Delta\mu_2$ can be computed as

$$\Delta\mu_1 = -\left(\mu_1 + \left(\frac{1}{t}\right)\mathrm{diag}(f_1)^{-1}\mathbf{1} + J_1^{-1} d\nu\right)$$
$$\Delta\mu_2 = -\left(\mu_2 + \left(\frac{1}{t}\right)\mathrm{diag}(f_2)^{-1}\mathbf{1} - J_2^{-1} d\nu\right)$$

Algorithm

Repeat

1. Compute $\Delta\nu, \Delta\mu_1, \Delta\mu_2$ by primal-dual interior point method
2. Backtracking line search with $\alpha = 0.01, \beta = 0.5$
3. Update $t = \max\left(\frac{2(n-2)\mu}{p-d}, 1.2t\right)$ where μ is the parameter for update t
4. Compute

   $p_1 = \frac{1}{2}(Dy - \mu_1 + \mu_2)^T(DD^T)^{-1}(Dy - \mu_1 + \mu_2) + \lambda(\mu_1 + \mu_2)$

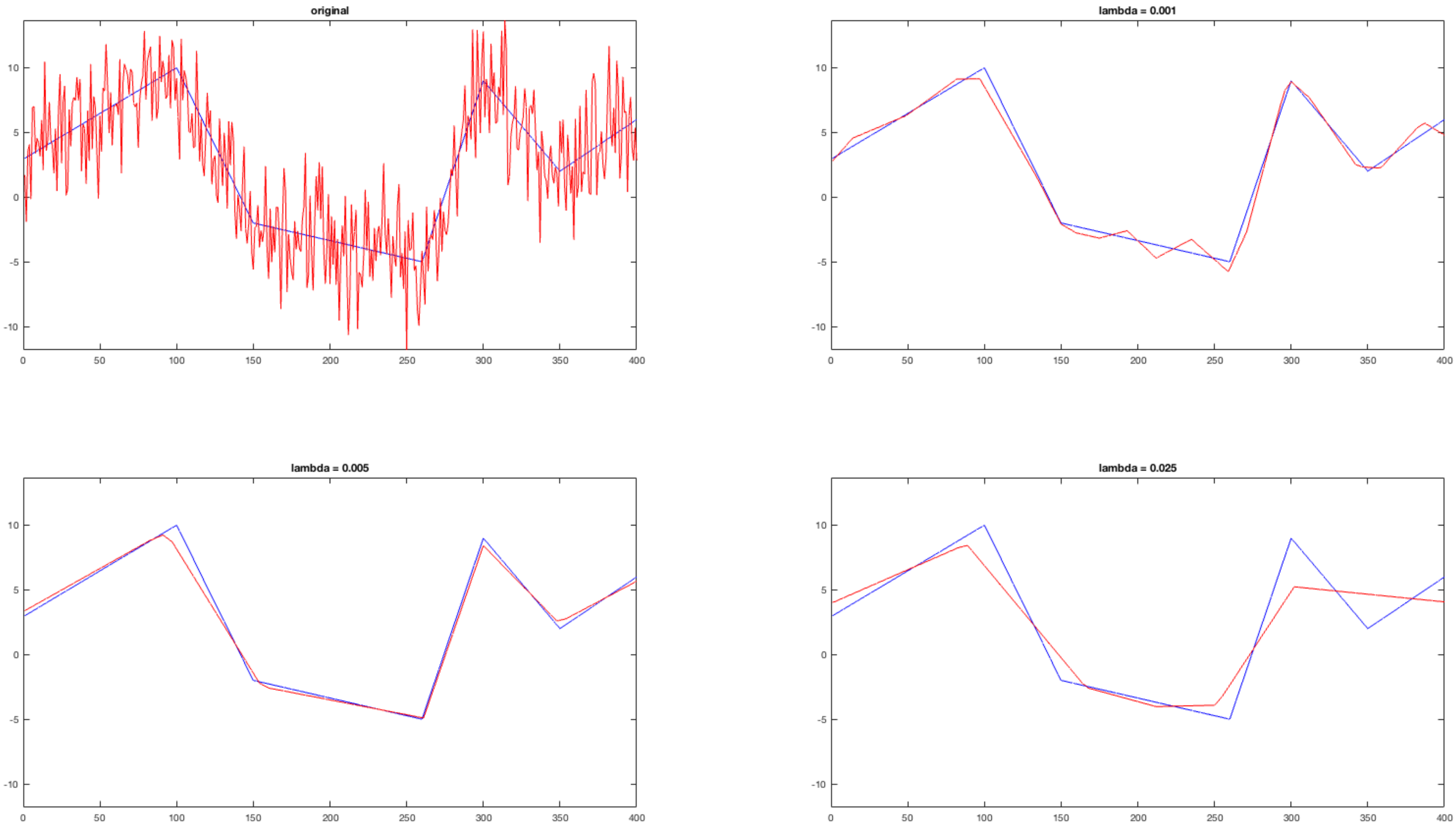
   $p_2 = \frac{1}{2}\nu^T DD^T \nu + \lambda\|Dy - DD^T\nu\|_1$

   $p = \max(p_1, p_2)$

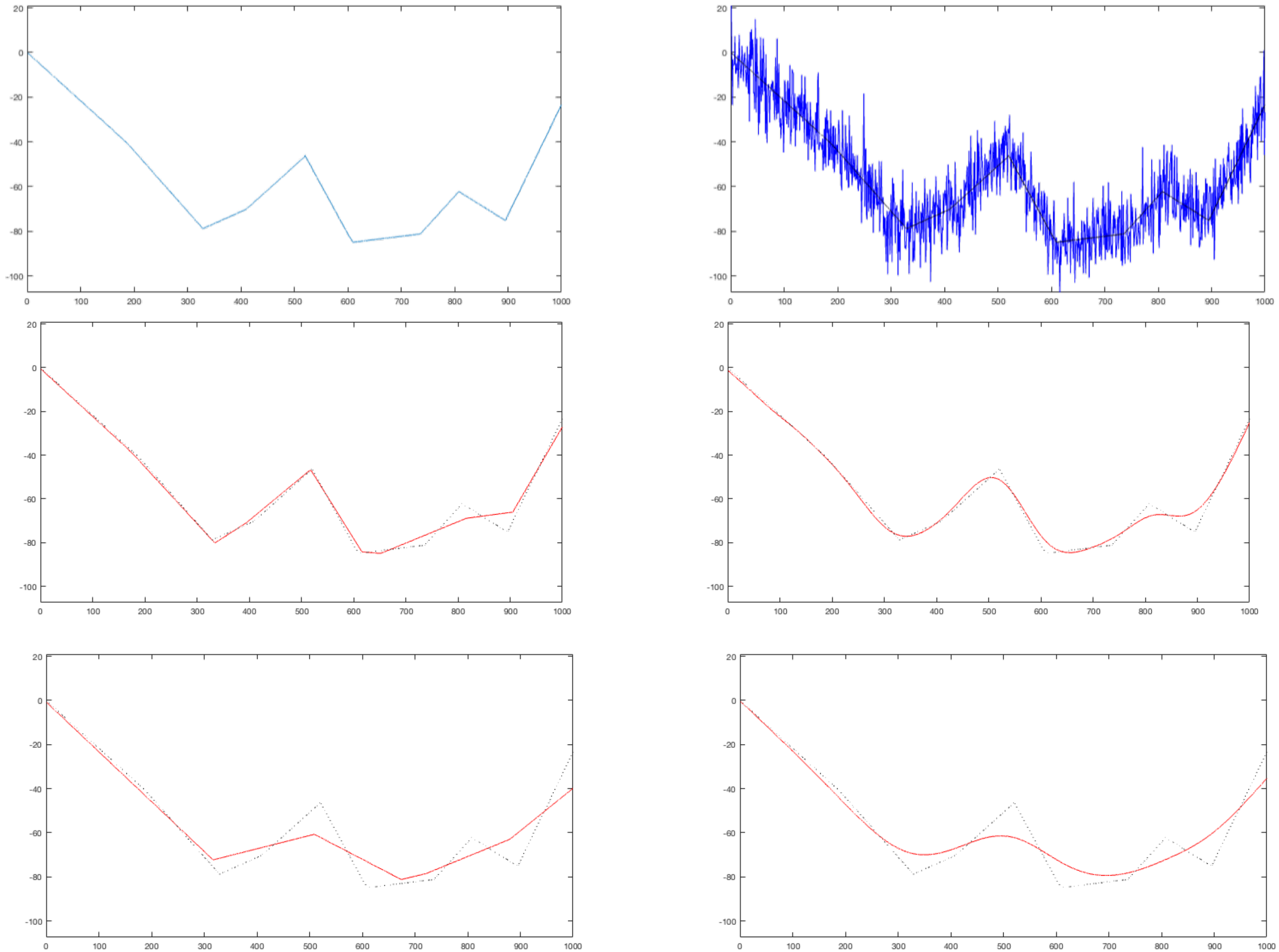
   $d = -\frac{1}{2}\nu^T DD^T \nu + y^T D\nu$

Until $p - d < \epsilon$

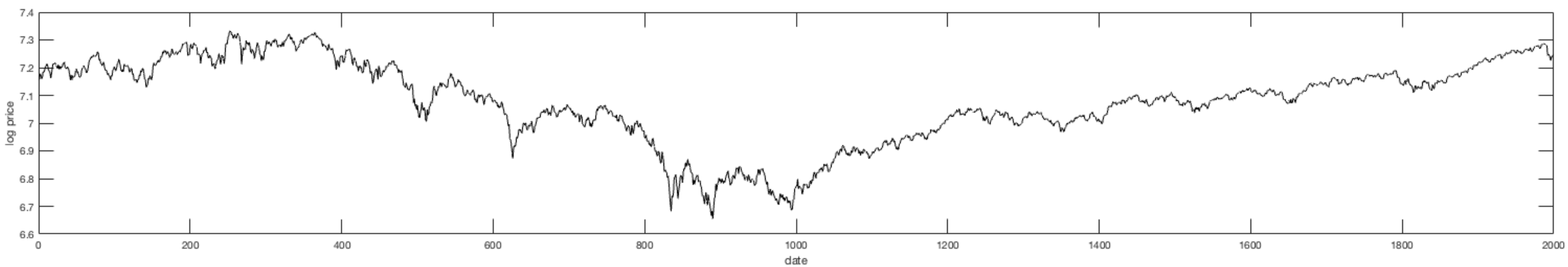
Trend estimate and regularization parameter λ

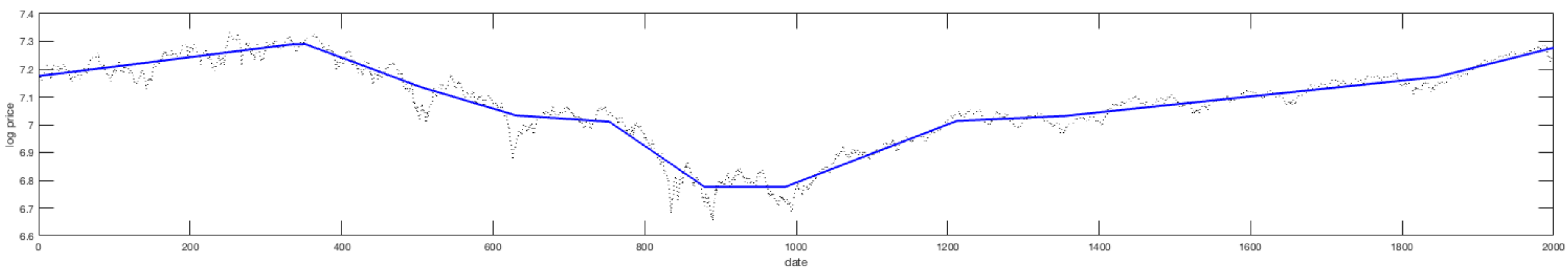


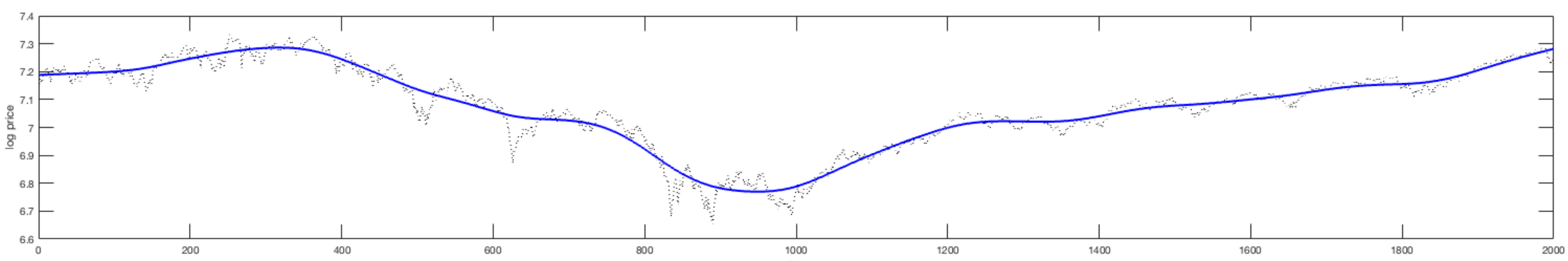
l_1 trend filter and H-P filter



Daily closing values of the S&P 500 Index







Application to robotics