Ville Väänänen

Maximum likelihood parameter estimation in discrete-time state-space models

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 14.3.2011

Thesis supervisor:

Prof. Jouko Lampinen

Thesis instructor:

D.Sc. (Tech.) Simo Särkkä

Aalto University
School of Electrical Engineering

Contents

1 Introduction

This thesis is about parameter estimation in dynamical systems...

2 State-space models

In this report we will be focusing on models that are defined in so called *state space* form. The state is denoted with $\mathbf{x}$, and it belongs to the space $\mathbb{R}^d$. More specifically the models we are considering can be written as

These equations can be understood as specifying a model, where at every step k the state $\mathbf{x}_k$ is observed by a noisy observation (or measurement) $\mathbf{y}_k$. Furthermore the states evolve as specified by the equation (1a). Thus the states form a *Markov chain*.

2.1 Linear

More specifically the models we are considering can be written as

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{q}_{k-1} \tag{1a}$$
$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_\mathbf{k} + \mathbf{r}_k \tag{1b}$$
$$\mathbf{q}_{k-1} \sim \mathrm{N}(0, \mathbf{Q}) \tag{1c}$$
$$\mathbf{r}_k \sim \mathrm{N}(0, \mathbf{R}) \tag{1d}$$
$$\mathbf{x}_0 \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{1e}$$

Equations (1a), (1b), (1c) and (1d) together specify the following conditional distributions

$$\mathbf{x}_k|\mathbf{x}_{k-1} \sim \mathrm{N}(\mathbf{A}\mathbf{x}_{k-1}, \mathbf{Q}) \tag{2}$$
$$\mathbf{y}_k|\mathbf{x}_k \sim \mathrm{N}(\mathbf{H}\mathbf{x}_\mathbf{k}, \mathbf{R}) \tag{3}$$

Linearity in this case means that $\mathbf{x}_k$ is a linear combination of the elements of $\mathbf{x}_{k-1}$ and $\mathbf{y}_k$ is a linear combination of the elements of $\mathbf{x}_k$ (with additive noise in both cases). Since the noise terms $\mathbf{q}_{k-1}$ and $\mathbf{r}_k$ are assumed to be white and Gaussian, these models are called linear-Gaussian.

2.2 Nonlinear

The SSM model is now

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{q}_{k-1} \tag{4a}$$
$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_\mathbf{k}) + \mathbf{r}_k \tag{4b}$$
$$\mathbf{q}_{k-1} \sim \mathrm{N}(0, \mathbf{Q}) \tag{4c}$$
$$\mathbf{r}_k \sim \mathrm{N}(0, \mathbf{R}) \tag{4d}$$
$$\mathbf{x}_0 \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{4e}$$

We assume an implicit dependence of f and h on the parameter $\boldsymbol{\theta}$.

2.3 Linear in the parameters

Suppose the function $\mathbf{f}$ is linear in the parameters and the dimension of the state $\mathbf{x}$ is d. Then in the most general case $\mathbf{f}(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}^d$ is a linear combination of vector valued functions $\boldsymbol{\rho}_k(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}^{d_k}$ and the parameters are matrices $\boldsymbol{\Phi}_k \in \mathbb{R}^{d \times d_k}$. More specifically, we have

$$\begin{aligned} \mathbf{f}(\mathbf{x}) &= \boldsymbol{\Phi}\boldsymbol{\rho}_1(\mathbf{x}) + \cdots + \boldsymbol{\Phi}_m\boldsymbol{\rho}_m(\mathbf{x}) \\ &= \begin{bmatrix} \boldsymbol{\Phi}_1 & \dots & \boldsymbol{\Phi}_m \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho}_1(\mathbf{x}) \\ \vdots \\ \boldsymbol{\rho}_m(\mathbf{x}) \end{bmatrix} \\ &= \mathbf{A}\mathbf{g}(\mathbf{x}), \end{aligned} \tag{5}$$

where $\mathbf{A} \in \mathbb{R}^{d \times \sum_{k=1}^m d_k}$ and $\mathbf{g}(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}^{\sum_{k=1}^m d_k}$. For example, in case of the function

$$f(x,t) = ax + b\frac{x}{1+x^2} + c\cos(1.2t) \tag{6}$$

we would have

$$f(x,t) = \begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} x \\ \frac{x}{1+x^2} \\ \cos(1.2t) \end{bmatrix} \tag{7}$$

Suppose now, that the matrix A depends on parameters $\mathbf{s}$.

3 Optimal filtering and smoothing

3.1 Kalman filtering

3.2 RTS smoothing

3.3 Example

3.4 Sigma point methods

3.4.1 Unscented Kalman filter

3.4.2 Gauss-Hermite Kalman filter

3.4.3 Cubature Kalman filter

3.5 Example

3.6 Particle filtering

4 Parameter estimation

5 Nonlinear optimization

5.1 Marginal likelihood

5.2 Gradient

6 Expectation Maximization