# Adjoint Gradients for System Identification

John Kormylo

Our objective is to identify certain parameters which describe a state vector model of the form

$$\mathbf{x}(k+1) = \Phi_\theta \mathbf{x}(k) + \boldsymbol{\gamma}_\theta \mu(k) \tag{1}$$
$$z(k) = \mathbf{h}'_\theta \mathbf{x}(k) + n(k) \tag{2}$$

where $\mu(k)$ and $n(k)$ are zero mean white Gaussian noise processes with variances

$$E\{\mu^2(k)\} = Q_\theta(k) \tag{3}$$

and

$$E\{n^2(k)\} = R_\theta(k) \tag{4}$$

and where $\boldsymbol{\theta}$ is a vector containing the desired system parameters. The objective function for system identification (negative log likelihood) can be expressed in terms of Kalman filter components as

$$J(\boldsymbol{\theta}) = \sum_{k=1}^{N} \frac{\tilde{z}_\theta^2(k|k-1)}{\eta_\theta(k)} + \ln\big[\eta_\theta(k)\big] \tag{5}$$

where $\tilde{z}_\theta(k|k-1)$ is the innovations process with modeled variance $\eta_\theta(k)$.

One can compute the gradient $dJ/d\boldsymbol{\theta}$ in a simple two pass algorithm by using Lagrange multipliers. The objective function is now given as

$$\begin{aligned} J(\boldsymbol{\theta}) = \sum_{k=1}^{N} \Bigg\{ & \frac{\tilde{z}^2(k|k-1)}{\eta(k)} + \ln\big[\eta(k)\big] \\ + & \mathbf{a}'(k|k-1)\big[\hat{\mathbf{x}}(k|k-1) - \Phi_\theta \hat{\mathbf{x}}(k-1|k-1)\big] \\ + & \mathbf{a}'(k|k)\big[\hat{\mathbf{x}}(k|k) - \hat{\mathbf{x}}(k|k-1) - \mathbf{K}(k)\tilde{z}(k|k-1)\big] \\ + & \mathrm{tr}\big\{B(k|k-1)\big[P(k|k-1) - \Phi_\theta P(k-1|k-1)\Phi'_\theta - \boldsymbol{\gamma}_\theta Q_\theta(k)\boldsymbol{\gamma}'_\theta\big]\big\} \\ + & \mathrm{tr}\Big\{B(k|k)\Big[P(k|k) - \big[I - \mathbf{K}(k)\mathbf{h}'_\theta\big]P(k|k-1)\Big]\Big\} \\ + & c(k)\big[\tilde{z}(k|k-1) - z(k) + \mathbf{h}'_\theta \hat{\mathbf{x}}(k|k-1)\big] \\ + & d(k)\big[\eta(k) - \mathbf{h}'_\theta P(k|k-1)\mathbf{h}_\theta - R_\theta(k)\big] \\ + & \mathbf{e}'(k)\Big[\mathbf{K}(k) - \frac{P(k|k-1)\mathbf{h}_\theta}{\eta(k)}\Big]\Bigg\} \end{aligned} \tag{6}$$

where $\mathbf{a}(k|k-1)$, $\mathbf{a}(k|k)$, $B(k|k-1)$, $B(k|k)$, $c(k)$, $d(k)$ and $\mathbf{e}(k)$ for $k = 1, 2, \ldots, N$ are all Lagrange multipliers. Minimizing this objective function with respect to the Lagrange multipliers will produce the Kalman filter equations. Note that we have dropped the dependence of $\tilde{z}(k|k-1)$ and $\eta(k)$ on parameter vector $\boldsymbol{\theta}$.

Minimizing this objective function w.r.t. $\mathbf{a}(k|k-1)$ yields

$$\hat{\mathbf{x}}(k|k-1) = \Phi_\theta \hat{\mathbf{x}}(k-1|k-1) \tag{7}$$

and w.r.t. $\mathbf{a}(k|k)$ yields

$$\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k)\tilde{z}(k|k-1) \tag{8}$$

and w.r.t. $B(k|k-1)$ yields

$$P(k|k-1) = \Phi_\theta P(k-1|k-1)\Phi'_\theta + \boldsymbol{\gamma}_\theta Q_\theta(k)\boldsymbol{\gamma}'_\theta \tag{9}$$

and w.r.t. $B(k|k)$ yields

$$P(k|k) = \big[I - \mathbf{K}(k)\mathbf{h}'_\theta\big]P(k|k-1) \tag{10}$$

and w.r.t. $c(k)$ yields

$$\tilde{z}(k|k-1) = z(k) - \mathbf{h}'_\theta \hat{\mathbf{x}}(k|k-1) \tag{11}$$

and w.r.t. $d(k)$ yields

$$\eta(k) = \mathbf{h}'_\theta P(k|k-1)\mathbf{h}_\theta + R_\theta(k) \tag{12}$$

and w.r.t. $\mathbf{e}(k)$ yields

$$\mathbf{K}(k) = \frac{P(k|k-1)\mathbf{h}_\theta}{\eta^(k)} \tag{13}$$

for $k = 1, 2, \ldots, N$.

Minimizing this objective function w.r.t. $\tilde{z}(k|k-1)$ yields

$$c(k) = \mathbf{K}'(k)a(k|k) - 2\frac{\tilde{z}(k|k-1)}{\eta(k)} \quad \forall k = 1, 2, \ldots, N \tag{14}$$

and w.r.t $\hat{\mathbf{x}}(k|k-1)$ yields

$$\mathbf{a}(k|k-1) = \mathbf{a}(k|k) - \mathbf{h}'_\theta c(k) \quad \forall k = 1, 2, \ldots, N \tag{15}$$

and w.r.t $\hat{\mathbf{x}}(k|k)$ yields

$$\mathbf{a}(k|k) = \Phi'_\theta \mathbf{a}(k+1|k) \quad \forall k = 1, 2, \ldots, N-1 \tag{16}$$

and for $k = N$ yields

$$\mathbf{a}(N|N) = \mathbf{0} \quad . \tag{17}$$

Minimizing this objective function w.r.t. $\mathbf{K}(k)$ yields

$$\begin{aligned} \mathbf{e}(k) &= \tilde{z}(k|k-1)\mathbf{a}(k|k) - B(k|k)P(k|k-1)\mathbf{h}_\theta \\ &= \eta(k)\left[\frac{\tilde{z}(k|k-1)}{\eta(k)}\mathbf{a}(k|k) - B(k|k)\mathbf{K}(k)\right] \quad \forall k = 1, 2, \ldots, N \end{aligned} \tag{18}$$

and w.r.t $\eta(k)$ yields

$$\begin{aligned} d(k) &= \frac{\tilde{z}^2(k|k-1)}{\eta^2(k)} - \frac{1}{\eta(k)} - \frac{\mathbf{e}'(k)P(k|k-1)\mathbf{h}_\theta}{\eta^2(k)} \\ &= \left[\frac{\tilde{z}(k|k-1)}{\eta(k)}\right]^2 - \frac{1 + \mathbf{K}'(k)\mathbf{e}(k)}{\eta(k)} \quad \forall k = 1, 2, \ldots, N \end{aligned} \tag{19}$$

and w.r.t $P(k|k-1)$ yields

$$B(k|k-1) = B(k|k)\big[I - \mathbf{K}(k)\mathbf{h}'_\theta\big] + \mathbf{h}_\theta d(k)\mathbf{h}'_\theta + \frac{\mathbf{h}_\theta \mathbf{e}'(k)}{\eta(k)} \quad \forall k = 1, 2, \ldots, N \tag{20}$$

and w.r.t $P(k|k)$ yields

$$B(k|k) = \Phi'_\theta B(k+1|k)\Phi_\theta \quad \forall k = 1, 2, \ldots, N-1 \tag{21}$$

and for $k = N$ yields

$$B(N|N) = 0 \quad . \tag{22}$$

Observe that these equations can be computed recursively in the anti-causal direction starting from (17) and (22). The only quantities which need to be stored between the forward and reverse passes are $\tilde{z}(k|k-1)$, $\eta(k)$ and $\mathbf{K}(k)$.

Returning to (6), the partial with respect to the $j^{\text{th}}$ component of $\boldsymbol{\theta}$ is given by

$$
\begin{aligned}
\frac{\partial J}{\partial \theta_j} = \sum_{k=1}^{N} \Bigg\{ \quad & - \quad \mathbf{a}'(k|k-1)\frac{\partial \Phi}{\partial \theta_j}\hat{\mathbf{x}}(k-1|k-1) \\
& - \quad \mathrm{tr}\Bigg\{ B(k|k-1)\left[\frac{\partial \Phi}{\partial \theta_j}P(k-1|k-1)\Phi' + \Phi' P(k-1|k-1)\frac{\partial \Phi}{\partial \theta_j}'\right]\Bigg\} \\
& - \quad \frac{\partial Q(k)\boldsymbol{\gamma}'}{\partial \theta_j}B(k|k-1)\boldsymbol{\gamma} \\
& - \quad Q(k)\boldsymbol{\gamma}' B(k|k-1)\frac{\partial \boldsymbol{\gamma}}{\partial \theta_j} \\
& + \quad \frac{\partial \mathbf{h}}{\partial \theta_j}' P(k|k-1)B(k|k)\mathbf{K}(k) \\
& + \quad c(k)\frac{\partial \mathbf{h}}{\partial \theta_j}'\hat{\mathbf{x}}(k|k-1) \\
& - \quad d(k)\left[2\mathbf{h}' P(k|k-1)\frac{\partial \mathbf{h}}{\partial \theta_j} + \frac{\partial R(k)}{\partial \theta_j}\right]\Bigg\}
\end{aligned} \tag{23}
$$

since $P(k|k-1)$ is symmetrical. Obviously computing the gradients themselve may require some additional informaton stored between passes, depending on the exact nature of $\boldsymbol{\theta}$.

# Appendix

Some Useful Matrix Identities

$$
\begin{aligned}
\mathrm{tr}\{AB\} &= \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}b_{ji} \\
\mathrm{tr}\{A\mathbf{u}\mathbf{v}'\} &= \mathbf{v}' A\mathbf{u} \\
\frac{\partial}{\partial A}\mathrm{tr}\{AB\} &= B' \\
\frac{\partial}{\partial B}\mathrm{tr}\{AB\} &= A' \\
\frac{\partial}{\partial B}\mathrm{tr}\{ABC\} &= (CA)' = A'C'
\end{aligned}
$$

## References:


R. L. Kashyap, *Maximum Likelihood Identification of Stochastic Linear Systems*, **IEEE Trans. on Automatic Control**, vol. AC-15, no. 1, pp. 25-34, Feb. 1970.

N. R. Sandel, Jr. & K. I. Yared, **Maximum Likelihood Identification of State Space Models for Linear Dynamic Systems**, MIT Technical Report ESL-R-814, 1978.

John Kormylo, **Maximum-Likelihood Seismic Deconvolution**, Ph.D. Dissertation, Dept. of Electrical Engineering, University of Southern California, 1979.