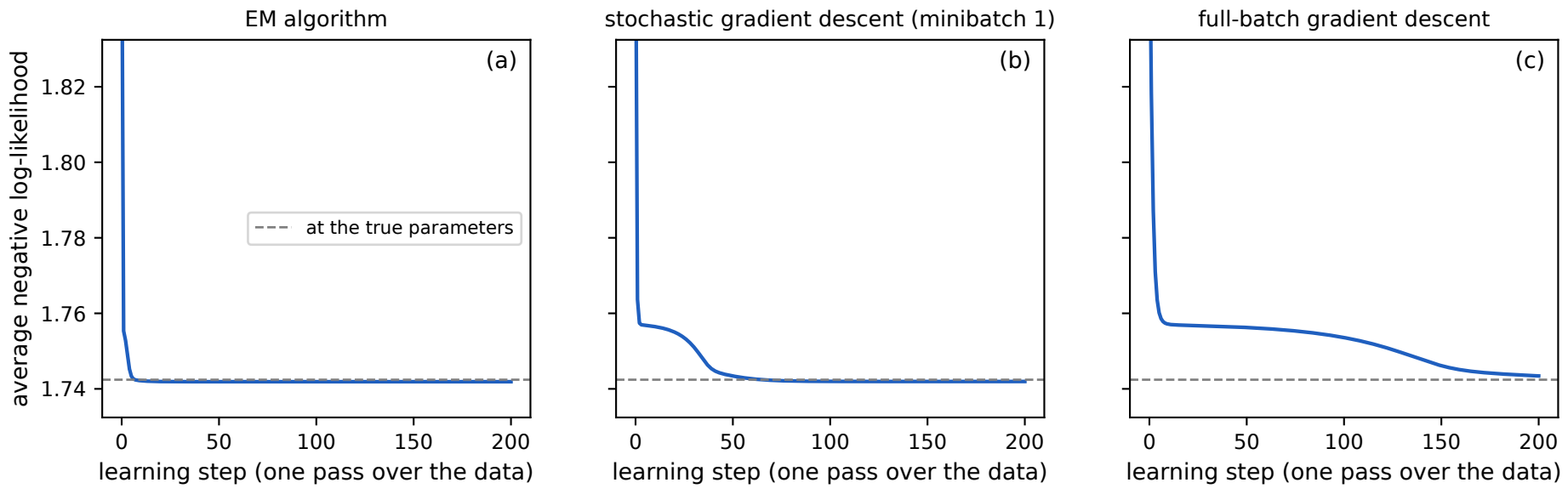

EM algorithm
stochastic gradient descent (minibatch 1)
full-batch gradient descent
(a)
(b)
(c)
average negative log-likelihood
1.82
1.80
1.78
1.76
1.74
at the true parameters
0
50
100
150
200
learning step (one pass over the data)