

Analytical Results for Dynamic Rational Inattention Problems

Bartosz Maćkowiak Filip Matějka
European Central Bank CERGE-EI

Mirko Wiederholt
Goethe University Frankfurt

February 2015

Abstract

Dynamic rational inattention problems are notoriously difficult to solve. We consider a standard linear quadratic Gaussian tracking problem, as in Sims (2003), Section 4. Let X_t denote the variable being tracked and let ε_t denote the time t innovation in the variable being tracked. We show that if the variable being tracked follows an AR(p) process, the optimal signal is about a linear combination of $\{X_t, \dots, X_{t-p+1}\}$ only. If the variable being tracked follows an ARMA(p,q) process, the optimal signal is about a linear combination of $\{X_t, \dots, X_{t-p+1}\}$ and $\{\varepsilon_t, \dots, \varepsilon_{t-q+1}\}$ only. Furthermore, the agent can attain the optimum with a single informative signal. These results make it much easier to solve dynamic rational inattention problems.

Consider the following rational inattention problem. The agent tracks X_t which follows a stationary Gaussian ARMA(p, q) process with finite $p \geq 1, q \geq 0$:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \theta_0 \varepsilon_t + \dots + \theta_q \varepsilon_{t-q}, \quad (1)$$

where ε_t follows a Gaussian white noise process with unit variance.

At each time $t \geq 1$, the agent receives N signals, where each signal is about a potentially different linear combination of current and past X_t and current and past ε_t . Formally,

$$S_t^N = \Theta X_t^M + \Phi \varepsilon_t^M + \psi_t^N, \quad (2)$$

where $X_t^M = (X_t, \dots, X_{t-M+1})'$, $\varepsilon_t^M = (\varepsilon_t, \dots, \varepsilon_{t-M+1})'$, and $\psi_t^N = (\psi_{t,1}, \dots, \psi_{t,N})'$. Here $M \geq \max\{p+1, q+1\}$ and $N \geq 1$. The noise components $\psi_{t,1}, \dots, \psi_{t,N}$ follow independent Gaussian white noise processes with vector of variances $\Upsilon = (\sigma_1^2, \dots, \sigma_N^2)$, and $\Theta, \Phi \in \mathbb{R}^{N \times M}$ are matrices of coefficients. The agent's information set at any time $t \geq 1$ includes any initial information and all signals received up to time t

$$\mathcal{I}_t = \mathcal{I}_0 \cup \{S_1^N, \dots, S_t^N\}. \quad (3)$$

The agent's decision problem reads

$$\min_{\Theta, \Phi \in \mathbb{R}^{N \times M}, \Upsilon \in \mathbb{R}_+^N} E \left[(X_t - E[X_t | \mathcal{I}_t])^2 \right], \quad (4)$$

subject to (1), (2), (3), and the constraint on information flow

$$\lim_{T \rightarrow \infty} \frac{1}{T} I(X_1, \dots, X_T; S_1^N, \dots, S_T^N) \leq \kappa. \quad (5)$$

The agent chooses the properties of the signals so as to minimize the mean square error subject to the information flow constraint. The expectations in (4) are computed using the steady-state Kalman filter.

Proposition 1 *Any solution to the agent's decision problem satisfies:*

$$\begin{aligned} \forall i &> p : \Theta_i = 0, \\ \forall i &> q : \Phi_i = 0, \end{aligned}$$

where Θ_i denotes the i th column of Θ and Φ_i denotes the i th column of Φ . This means, for $q = 0$, all informative signals are on linear combinations of $\{X_t, \dots, X_{t-p+1}\}$ only. For $q > 0$, all informative signals are on linear combinations of $\{X_t, \dots, X_{t-p+1}\}$ and $\{\varepsilon_t, \dots, \varepsilon_{t-q+1}\}$ only. Moreover, the agent can attain the optimum with a single informative signal.

The next pages give the proof in the case of $q = 0$ and $\Phi = 0$.

Proof. First, we show that for any signal of the form (2), the information flow constraint (5) is equivalent to

$$\lim_{T \rightarrow \infty} [H(X_T, \dots, X_{T-M+1} | S^{N,T-1}) - H(X_T, \dots, X_{T-M+1} | S^{N,T})] \leq \kappa. \quad (6)$$

Let

$$X^t = \{X_1, \dots, X_t\}, \quad S^{N,t} = \{S_1^N, \dots, S_t^N\}.$$

It follows from the chain rule for entropy that, $\forall T \geq M$,

$$H(S^{N,T}) = H(S^{N,M-1}) + \sum_{t=M}^T H(S_t^N | S^{N,t-1}),$$

and

$$H(S^{N,T} | X^T) = H(S^{N,M-1} | X^T) + \sum_{t=M}^T H(S_t^N | S^{N,t-1}, X^T).$$

Since conditioning reduces entropy, $\forall t \in \{M, \dots, T\}$,

$$H(S_t^N | S^{N,t-1}, X^T) \leq H(S_t^N | S^{N,t-1}, X_t, \dots, X_{t-M+1}).$$

For any signal of the form (2), this weak inequality holds with equality because both sides of the weak inequality simply equal $H(\psi_t^N)$. Collecting results yields that, $\forall T \geq M$, the mutual information between X^T and $S^{N,T}$ is given by

$$\begin{aligned} I(X^T; S^{N,T}) &= H(S^{N,T}) - H(S^{N,T} | X^T) \\ &= H(S^{N,M-1}) - H(S^{N,M-1} | X^T) \\ &\quad + \sum_{t=M}^T [H(S_t^N | S^{N,t-1}) - H(S_t^N | S^{N,t-1}, X_t, \dots, X_{t-M+1})]. \end{aligned}$$

It follows from the symmetry of mutual information that

$$\begin{aligned} &H(S_t^N | S^{N,t-1}) - H(S_t^N | S^{N,t-1}, X_t, \dots, X_{t-M+1}) \\ &= I(X_t, \dots, X_{t-M+1}; S_t^N | S^{N,t-1}) \\ &= H(X_t, \dots, X_{t-M+1} | S^{N,t-1}) - H(X_t, \dots, X_{t-M+1} | S^{N,t-1}, S_t^N). \end{aligned}$$

Combining results yields that, $\forall T \geq M$,

$$\begin{aligned} &I(X^T; S^{N,T}) \\ &= H(S^{N,M-1}) - H(S^{N,M-1} | X^T) \\ &\quad + \sum_{t=M}^T [H(X_t, \dots, X_{t-M+1} | S^{N,t-1}) - H(X_t, \dots, X_{t-M+1} | S^{N,t})]. \end{aligned}$$

We show below that the following limit exists

$$\lim_{T \rightarrow \infty} [H(X_T, \dots, X_{T-M+1} | S^{N,T-1}) - H(X_T, \dots, X_{T-M+1} | S^{N,T})]. \quad (7)$$

The last equation and Cesaro mean then imply

$$\begin{aligned} &\lim_{T \rightarrow \infty} \frac{1}{T} I(X^T; S^{N,T}) \\ &= \lim_{T \rightarrow \infty} [H(X_T, \dots, X_{T-M+1} | S^{N,T-1}) - H(X_T, \dots, X_{T-M+1} | S^{N,T})]. \end{aligned}$$

Second, one can express the information flow constraint (6) in terms of two conditional variance-covariance matrices, which can be computed from the steady-state Kalman filter equations. Let $\Sigma_{t|t}$ denote the conditional variance-covariance matrix of X_t^M given $S^{N,t}$. Let $\Sigma_{t|t-1}$ denote the conditional variance-covariance matrix of X_t^M given $S^{N,t-1}$. The signal has the following state-space representation

$$\begin{aligned} X_{t+1}^M &= F X_t^M + \begin{pmatrix} \theta_0 \varepsilon_{t+1} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \\ S_t^N &= \Theta X_t^M + \psi_t^N. \end{aligned}$$

The matrix F is a $(M \times M)$ -matrix given by

$$F_{ij} = \begin{cases} \phi_j & \text{if } i = 1 \text{ and } j \in \{1, \dots, p\} \\ 1 & \text{if } i = j + 1 \\ 0 & \text{otherwise} \end{cases}.$$

Recall that X_t follows a stationary process. It follows from Propositions 13.1-13.2 in Hamilton (1994) that the limits $\lim_{t \rightarrow \infty} \Sigma_{t|t}$ and $\lim_{t \rightarrow \infty} \Sigma_{t|t-1}$ exist and are given by

$$\begin{aligned} \Sigma_1 &= F \left[\Sigma_1 - \Sigma_1 \Theta' (\Theta \Sigma_1 \Theta' + R)^{-1} \Theta \Sigma_1 \right] F' + Q \\ \Sigma_0 &= \Sigma_1 - \Sigma_1 \Theta' (\Theta \Sigma_1 \Theta' + R)^{-1} \Theta \Sigma_1, \end{aligned}$$

where $\Sigma_1 \equiv \lim_{t \rightarrow \infty} \Sigma_{t|t-1}$, $\Sigma_0 \equiv \lim_{t \rightarrow \infty} \Sigma_{t|t}$, and Q and R are the variance-covariance matrix of the innovation in the state equation and the variance-covariance matrix of the innovation in the observation equation, respectively. Finally, conditional normality implies

$$\begin{aligned} H(X_t, \dots, X_{t-M+1} | S^{N,t-1}) &= \frac{1}{2} \log_2 \left[(2\pi e)^M \det \Sigma_{t|t-1} \right], \\ H(X_t, \dots, X_{t-M+1} | S^{N,t}) &= \frac{1}{2} \log_2 \left[(2\pi e)^M \det \Sigma_{t|t} \right]. \end{aligned}$$

Hence, the limit (7) exists and equals $\frac{1}{2} \log_2 \left(\frac{\det \Sigma_1}{\det \Sigma_0} \right)$. The information flow constraint (6) can be expressed as

$$\frac{1}{2} \log_2 \left(\frac{\det \Sigma_1}{\det \Sigma_0} \right) \leq \kappa. \quad (8)$$

Third, if we have two signals $\{\Theta, \Upsilon\}$ that yield the same upper-left $(p \times p)$ sub-matrix of Σ_0 and one signal has the property $\Theta_i = 0$ for all $i > p$, while the other signal does not have this property, then the first signal yields a lower value of the left-hand side of the information flow constraint (6) even though

both signals imply the same value of the objective (4). To see this, note that the information flow constraint (6) is equivalent to

$$\lim_{T \rightarrow \infty} \begin{pmatrix} H(X_T, \dots, X_{T-p+1} | S^{N,T-1}) - H(X_T, \dots, X_{T-p+1} | S^{N,T}) \\ + H(X_{T-p}, \dots, X_{T-M+1} | S^{N,T-1}, X_T, \dots, X_{T-p+1}) \\ - H(X_{T-p}, \dots, X_{T-M+1} | S^{N,T}, X_T, \dots, X_{T-p+1}) \end{pmatrix} \leq \kappa.$$

Furthermore,

$$\begin{aligned} & H(X_{T-p}, \dots, X_{T-M+1} | S^{N,T-1}, X_T, \dots, X_{T-p+1}) \\ & - H(X_{T-p}, \dots, X_{T-M+1} | S^{N,T}, X_T, \dots, X_{T-p+1}) \\ = & H(X_{T-p}, \dots, X_{T-M+1} | S^{N,T-1}, X_T, \dots, X_{T-p+1}) \\ & - H(X_{T-p}, \dots, X_{T-M+1} | S^{N,T-1}, S_T^N, X_T, \dots, X_{T-p+1}) \\ = & H(X_{T-p}, \dots, X_{T-M+1} | S^{N,T-1}, X_T, \dots, X_{T-p+1}) \\ & - H(X_{T-p}, \dots, X_{T-M+1} | S^{N,T-1}, \Theta \hat{X}_T^M + \psi_T^N, X_T, \dots, X_{T-p+1}), \end{aligned}$$

where $\hat{X}_T^M = (0, \dots, 0, X_{T-p}, \dots, X_{T-M+1})'$. The last difference is non-negative, since conditioning weakly reduces entropy, and it is zero if and only if $\Theta_i = 0$ for all $i > p$. In addition, two signals $\{\Theta, \Upsilon\}$ that yield the same upper-left $(p \times p)$ sub-matrix of Σ_0 also generate the same upper-left $(p \times p)$ sub-matrix of Σ_1 . This is because X_t follows an AR(p) process. Hence, two signals $\{\Theta, \Upsilon\}$ that yield the same upper-left $(p \times p)$ sub-matrix of Σ_0 generate the same limit

$$\lim_{T \rightarrow \infty} [H(X_T, \dots, X_{T-p+1} | S^{N,T-1}) - H(X_T, \dots, X_{T-p+1} | S^{N,T})] \leq \kappa.$$

Finally, two signals $\{\Theta, \Upsilon\}$ that yield the same upper-left $(p \times p)$ sub-matrix of Σ_0 imply the same value of the objective (4) because the objective is the upper-left entry of Σ_0 .

Fourth, we show that for any signal $\{\hat{\Theta}, \hat{\Upsilon}\}$ with $\hat{\Theta}_i \neq 0$ for some $i > p$, there exists a signal $\{\Theta, \Upsilon\}$ with $\Theta_i = 0$ for all $i > p$ that yields the same upper-left $(p \times p)$ sub-matrix of Σ_0 .

For any variance-covariance matrices $\tilde{\Sigma}_1, \tilde{\Sigma}_0$ there exists a signal generating posterior uncertainty $\tilde{\Sigma}_0$ from the prior uncertainty $\tilde{\Sigma}_1$ if and only if $\tilde{\Sigma}_1 - \tilde{\Sigma}_0$ is positive-semi definite. This is what is in rational inattention often called the no-forgetting constraint.

Since Σ_0 is generated from Σ_1 by signal $\{\hat{\Theta}, \hat{\Upsilon}\}$, then $(\Sigma_0 - \Sigma_1)$ must be positive semi-definite. By Sylvester's criterion (Bazaraa et al., 2013), all the upper-left $(p \times p)$ sub-matrix of $(\Sigma_0 - \Sigma_1)$ are positive semi-definite, too. using the statement above, this implies that there exists a signal $\{\Theta, \Upsilon\}$ with $\Theta_i = 0$ for all $i > p$ that generates upper-left $(p \times p)$ sub-matrix of Σ_0 .

Therefore, for any $\{\hat{\Theta}, \hat{\Upsilon}\}$ with $\hat{\Theta}_i \neq 0$ for some $i > p$ we can find another signal $\{\Theta, \Upsilon\}$ generates the same objective at a lower information flow. This concludes the proof. ■

References

Bazaraa, M. S., Sherali, H. D., and Shetty, C. M. (2013). *Nonlinear programming: theory and algorithms*. John Wiley & Sons.