Using the Foundations and Trends® LaTeX Class

Instructions for Creating an FnT Article

Suggested Citation: Alet Heezemans and Mike Casey (2018), "Using the Foundations and Trends® LaTeX Class", : Vol. xx, No. xx, pp 1–18. DOI: 10.1561/XXXXXXXXX.

Alet Heezemans
now publishers, Inc.
alet.heezemans@nowpublishers.com

Mike Casey
now publishers, Inc.
mike.casey@nowpublishers.com

now
the essence of knowledge
Boston — Delft

Contents

Using the Foundations and Trends® LaTeX Class

Alet Heezemans[1] and Mike Casey[2]

[1]*now publishers, Inc.; alet.heezemans@nowpublishers.com*
[2]*now publishers, Inc.; mike.casey@nowpublishers.com*

ABSTRACT

This document describes how to prepare a Foundations and Trends® article in LaTeX . The accompanying LaTeX source file FnTarticle.tex (that produces this output) is an example of such a file.

Alet Heezemans and Mike Casey (2018), "Using the Foundations and Trends® LaTeX Class", : Vol. xx, No. xx, pp 1–18. DOI: 10.1561/XXXXXXXXX.

1

Introduction to Reinforcement Learning

The reinforcement learning (RL) method aims to make the sequential decision:

1.1 Sequential Decision Problem

Background Suppose $N > 0$ is the *time horizon* of the decision problem. The symbol $x_k \in \mathcal{X}_k$ refers to the *state* at time k, for each $k \in [0, N+1]$. At time k, after obverving the state x_k, we are required to take an appropriate action $a_k \in \mathcal{A}_k$. Given (x_k, a_k), a new (*random*) state x_{k+1} is observed and a *one-step* (*stage*) cost $g_k(x_k, a_k, x_{k+1})$ is incurred. The sequence $(x_0, a_0, x_1, a_1, \ldots, x_N, a_N, x_{N+1})$ is known as an *episode*.

Objective The goal is to minimize the expected total cost:

1. The total cost for one given episode is

$$\sum_{k=0}^{N} g_k(x_k, a_k, x_{k+1}) \tag{1.1}$$

2. Here $\pi = \{\mu_0, \ldots, \mu_N\}$ is called a policy, with a collection of functions
$$\mu_k : \mathcal{X}_k \to \mathcal{A}_k, \quad a_k = \mu_k(x_k) \tag{1.2}$$

3. The expected total cost under total cost π is given by:
$$J_\pi(x) = \mathbb{E}_\pi \left[\sum_{k=0}^{N} g_k(x_k, a_k, x_{k+1}) \middle| x_0 = x \right] \tag{1.3}$$
where $\mathbb{E}_\pi$ is the expectation over randomness for transitions from x_k to x_{k+1}, $k \in [0, N]$, under the policy π.

4. Therefore, we aim to solve the policy-based optimization problem given the initial state x:
$$J^*(x) = \inf_\pi J_\pi(x) \tag{1.4}$$

We say a policy π^* is an *optimal policy* if $J^*(x) = J_{\pi^*}(x)$. However, the infimum may not be achievable, i.e., the optimal poplicy may not exist. In our setting, we only focus on the case where the optimal policy can be achieved.

Possible Types of Dynamics The transition from x_k, a_k to x_{k+1} can be modelled by the transition probability:

$$\mathbb{P}\{x_{k+1} = y | x_k = x, a_k = a\} = P_k(x, a, y), \quad \forall x \in \mathcal{X}_k, a_k \in \mathcal{A}_k, y \in \mathcal{X}_{k+1}.$$

We will possibly face three different cases for this problem

1. When the transition probabilities are known, this is the so-called *perfect information* case. This problem reduces to the *markov decision problem*, which can be solved by applying *dynamic programming*

2. When the transition probabilities are unknown, but we can get many epsiodes from data. In this case, we can estimate the transition probability, but the computation cost is expansive.

3. When the transition probabilities are unknown, but given (x_k, a_k), we can sample for x_{k_1}, i.e., *a simulator is available*. This case is between case 1 and 2, and the corresponding solution will be talked later.

Solving for Case (1)

- When N is finite, We set the cost-to-go function J_k to be the total cost starting from the k-th stage into the final state:

$$\begin{aligned} J_{N+1} &= 0, \qquad x \in \mathcal{X}_{N+1}, \text{ and for } k = N, N-1, \ldots, 0, \\ J_k(x) &= \min_{a \in \mathcal{A}_k(x)} \mathbb{E}_{P_k(x,a,y)} \left[g_k(x,a,y) + J_{k+1}(y) \right], \quad x \in \mathcal{X}_k \\ &= \min_{a \in \mathcal{A}_k(x)} \sum_{y \in \mathcal{X}_{k+1}} P_k(x,a,y) \left[g_k(x,a,y) + J_{k+1}(y) \right] \\ J^*(x) &= J_0(x), \quad x \in \mathcal{X}_0 \end{aligned}$$

  This is so-called *backward induction algorithm*. The drawback is due to its huge complexity. To obtain the optimal solution, when N is large, we need $\prod_{k=0}^{N} |\mathcal{A}_k||\mathcal{X}_{k+1}|$ computation times, though in some special problem this complexity can be reduced.

- When N is infinite, the problem becomes easier to solve suprisingly. Assume the time homogeneity with a discounted factor $\beta < 1$, i.e., the transition probability P_k remains the same for large k, and the stage cost gradually decreases with discounted factor β and underlying cost function g:

$$P_k = P, \quad g_k = \beta^k g$$

  As a result, it suffices to solve the *Bellman equation*:

$$J^*(x) = \min_{a \in \mathcal{A}} \sum_{y \in \mathcal{X}} P(x,a,y) \left(g(x,a,y) + \beta J^*(y) \right) \tag{1.5}$$

Let's discuss the Bellman equation in detail.

1.2 Bellman's Equation

Here the key assumption follows from the last section: Assume the time homogeneity with a discounted factor $\beta < 1$.

Notation We use the notation $P(x,a,y)$ to denote the conditional probability $\mathbb{P}\{x' = y \mid x, a\}$ for $y \in \mathcal{X}$, and write the summation over

probabilties in an expectation form:

$$\sum_{x' \sim P(\cdot|x,a)} \left[g(x,a,x') + \beta J^*(x') \right]$$

Actually, Bellaman [Bertsekas, 1995] tells us that we can find the solution from the Bellaman equation:

Theorem 1.1. Assume that the state space $\mathcal{X}$ and action space $\mathcal{A}$ is finite. The optimal value function J^* is the *one and only one* solution solution to the Bellman's equation:

$$J^*(x) = \min_{a \in \mathcal{A}} \mathbb{E}_{x' \sim P(\cdot|x,a)} \left(g(x,a,x') + \beta J^*(x') \right) \tag{1.6}$$

After obtaining the optimal solution J^*, we can derive a policy J^*, which is one of the optimal policy to our decision problem:

Theorem 1.2. Assume that the state space $\mathcal{X}$ and action space $\mathcal{A}$ is finite. Let $J^* : \mathcal{X} \to \mathbb{R}$ be the unique solution to the Bellman equation. Define $\mu^* : \mathcal{X} \to \mathcal{A}$ via

$$\mu^*(x) = \arg \min_{a \in \mathcal{A}(x)} \mathbb{E}_{x' \sim P(\cdot|x,a)} \left(g(x,a,x') + \beta J^*(x') \right), \qquad \forall x \in \mathcal{X}$$

Notation

- For a function $h \in \mathcal{X} \to \mathbb{R}$, define the *h-greedy policy* $\mu_h : \mathcal{X} \to \mathcal{A}$ via

  $$\mu_h(x) = \arg \min_{a \in \mathcal{A}(x)} \mathbb{E}_{x' \sim P(\cdot|x,a)} \left(g(x,a,x') + \beta h(x') \right), \qquad \forall x \in \mathcal{X}$$

  Here you may understand h as the tradeoff for minimizing between the current cost g and the future cost, and therefore J^*-greedy policy tells us how to make optimal policy for the given decision problem.

- Also define the *Bellman operator* T for J: $\mathcal{X} \to \mathbb{R}$, with

  $$(TJ)(x) = \min_{a \in \mathcal{A}(x)} \mathbb{E}_{x' \sim P(\cdot|x,a)} \left(g(x,a,x') + \beta J(x') \right) \tag{1.7}$$

- Fix a stationary policy μ, we assume that $\mu(x)$ admits only one value (randomly choose one value from $\mu(x)$ is also ok), and therefore define the Bellman operator for J w.r.t. μ: $\mathcal{X} \to \mathbb{R}$, with

$$(T_\mu J)(x) = \mathbb{E}_{x' \sim P(\cdot|x,a)} \left(g(x, \mu(x), x') + \beta J(x') \right) \tag{1.8}$$

- Therefore for any J,

$$(T_{\mu_J} J)(x) = (TJ)(x), \ \forall x \in \mathcal{X} \tag{1.9}$$

Method for solving Bellman equation: value iteration Starting with arbitrary $J^0 := J$, we define the value iteration function $J^n = TJ^{n-1}$. The *value iteration* guarantees to converge finally:

Theorem 1.3. For any function $J : \mathcal{X} \to \mathbb{R}$, we have for all $x \in \mathcal{X}$:

$$J^*(x) = \lim_{N \to \infty} (T^N J)(x).$$

The convergence rate is easy to show:

$$\begin{aligned} \max_{x \in \mathcal{X}} |J^N(x) - J^*(x)| &= \max_{x \in \mathcal{X}} |TJ^{N-1}(x) - TJ^*(x)| \\ &\leq \beta \max_{x \in \mathcal{X}} |J^{N-1}(x) - J^*(x)| \\ &\leq \beta^N \max_{x \in \mathcal{X}} |J(x) - J^*(x)|. \end{aligned}$$

Therefore, we have two alternatives to guarantee fast convergence rate. One is to find a small discount factor; another way is to make good initial guess for J^0.

Complexity Analysis for value evaluation The calculation for $J^n(x) = (TJ^{n-1})(x)$ requires the complexity $|\mathcal{A}||\mathcal{X}|$:

$$J^n(x) = \min_{a \in \mathcal{A}(x)} \left[g(x, a) + \beta \sum_{x \in \mathcal{X}} P(x, a, y) J^{n-1}(y) \right]$$

Therefore the complexity for N steps is $N \cdot |\mathcal{A}| \cdot |\mathcal{X}|^2$.

1.2.1 Searching for optimal policy

Criteria for the performance of policy Given a stationary policy μ, assume that the *value function* J_μ satisfies the Bellman equation

$$J_\mu(x) = g(x, \mu(x)) + \beta \sum_{y \in \mathcal{X}} P(x, \mu(x), y) J_\mu(y), \qquad x \in \mathcal{X} \tag{1.10}$$

Therefore, the expected cost w.r.t. μ is given by:

$$J_\mu = (I - \beta P_\mu)^{-1} g_\mu, \tag{1.11}$$

where g_μ is an $|\mathcal{X}|$-vector with entries $g_\mu(x) = g_\mu(x, \mu(x))$; P_μ is an $|\mathcal{X}| \times |\mathcal{X}|$ matrix with entries $P_\mu(x, y) = P(x, \mu(x), y)$.

The complexity of (1.11) is huge, but there are many other algorithms for solving (or approximately solving) (1.10).

Algorithm for searching for optimal policy The basic algorithm for deriving an optimal (stationary) policy is listed as follows:

1. Initialization: guess an initial stationary policy μ^0.
2. Policy evaluation: find J_{μ^k}
3. Policy Improvement: Obtain a new stationary policy μ^{k+1} that is J_{μ^k}-greedy
4. Stop Criteria: if $J_{\mu^k} = J_{\mu^{k+1}}$, then stop, otherwise go to step 2.

$$\underset{\text{old policy}}{\mu^k} \longrightarrow \underset{\text{value function of the old policy}}{J_{\mu^k}} \longrightarrow \underset{\text{new improved policy}}{\mu^{k+1}}$$

Figure 1.1: Diagram for the algorithm above

Convergence Issue Bellman gives the convergence of the algorithm, given the condition that the value function can be *exactly evaluated*:

Theorem 1.4. Fix a policy μ. Let $\hat{\mu}$ be a J_μ-greedy policy, then we have

$$J_{\hat{\mu}}(x) \leq J_\mu(x), \text{ for each } x \in \mathcal{X}.$$

Moreover, if μ is not optimal, strict inequality holds for at least one state.

Remark 1.1. This theorem gives the convergence of the algorithm to the optimal policy. However, once we use algorithms to find the approximated value function due to the pity of hugh complexity for obtaining exact value function, the convergence to the optimal policy is no longer guaranteed

1.3 Reinforcement Programming

- The dynamic programming aims to solve the decision problem with the case where the *model of the environment's dynamics is given (P, g are known).*
- The reinforcement programming aims to solve the decision problem with the case where the *model of the environment's dynamics is not given (P, g are unknown), but the samples are available.*

Let's now focus on reinforcement learning only. First let's discuss issues about policy evaluation:

1.3.1 Policy Evaluation

Given a stationary policy μ, the definition for value function $J_\mu(x)$ is given by:

$$J_\mu(x) = \mathbb{E}\left[\sum_{k=0}^{\infty} \beta^k g(x_k, \mu(x_k), x_{k+1}) \middle| x_0 = x\right]$$

To simplify the computation complexity, we find that the J_μ has to satisfy the one-step Bellman equation:

$$\begin{aligned} J_\mu(x) &= (T_\mu J_\mu)(x) \\ &= g_\mu(x) + \beta(P_\mu J_\mu)(x) \\ &= \mathbb{E}\left[g(x, \mu(x), x_1)\right] + \beta \mathbb{E}[J_\mu(x_1)] \end{aligned}$$

To get the approximated value function, we can equivalently solve

$$J_\mu = \arg\min_{J\in\mathbb{R}^{|\mathcal{X}|}} \sum_{x\in\mathcal{X}} |J(x) - (T_\mu J)(x)|^2 \xi(x) \tag{1.12}$$

where the weight ξ could be anything *if (1.12) is solvable.*

Bellman Error Therefore, there is a popular way to define the quality of the approximated value function:

$$\|J - T_\mu J\|_\xi^2 \equiv \sum_{x\in\mathcal{X}} |J(x) - (g_\mu(x) + \beta(P_\mu J)(x))|^2 \xi(x) \tag{1.13}$$

where ξ is commonly chosen as the stationary distribution of the DTMC with transition matrix P_μ, i.e., more visited states should be weighted more. Note that the ξ may not be explicitly derived, but only for simplicity of analysis.

Optimization for the loss function (1.13) By taking gradient of (1.13) wr.t. J to 0, we obtain:

$$D(J - (g_\mu + \beta P_\mu J)) = 0, \tag{1.14}$$

where D is the diagonal matrix with ξ along the diagonal. Note that (1.14) is equivalent to solving the fixed point problem

$$J = J - \gamma \underbrace{D[(I - \beta P_\mu)J - g_\mu]}_{(*)}, \tag{1.15}$$

where the term (*) can be expressed as

$$D[(I - \beta P_\mu)J - g_\mu] = \mathbb{E}_\xi \left[J(x) - \beta J(x') - g(x, x')\right] \tag{1.16}$$

Tabular TD(0) Learning The Equations (1.13)-(1.16) motivates the TD learning below:

1. (Initialization) Initialize J^0, choose initial state x_0.
2. (Simulation) simulate one step (once) staring from x_k, with the decision given by μ. Our observation is the next state x_{k+1} and the one-step cost g_k

3. (Update) Our update is as follows:

$$\begin{cases} J^{k+1}(x_k) = J^k(x_k) - \gamma_k \left[J^k(x_k) - (g_k + \beta J^k(x_{k+1})) \right], \\ J^{k+1}(y) = J^k(y), \text{ for } y \neq x_k \end{cases} \tag{1.17}$$

4. Set $k = k + 1$ and go to step 2.

Remark 1.2. This algorithm only reuqires the update of parameters locally. The drawback for many other algorithms is that they require the global update for parameters, which is computationally difficult.

The literature [Sutton, 1988] gives the convergence analysis for this algorithm:

Theorem 1.5. Asume a Markov chain associated with the policy μ is *finite, irreducible, and aperiodic*. Given bounded costs $|g(x, a, x')| < G$ and learning rate such that, $\sum_{k=0}^{\infty} \gamma_k = \infty, \sum_{k=0}^{\infty} \gamma_k^2 < \infty$, we have

$$\lim_{k \to \infty} J^k = J_\mu \text{ a.s.}$$

Remark 1.3. One can have a similar algorithm but without focusing on one-step Bellman equation but ℓ-step:

$$J_\mu(x) = \mathbb{E}\left[\sum_{k=0}^{\infty} \beta^k g(x_k, \mu(x), x_{k+1}) + \beta^{\ell+1} J_\mu(x_{\ell+1}) \right]$$

where $\ell \sim \text{Geometric}(\lambda)$ for $0 \leq \lambda \leq 1$. The corresponding algorithm is TD(λ) learning algorithm, which will be discussed in the next lecture.

1.3.2 Policy Improvement

Then we turn into the policy improvement.

***Q*-Learning** Approximation for J_μ-greedy function requires defining an optimal Q-factor:

$$\begin{aligned} Q^*(x, a) &= \sum_{x' \in \mathcal{X}} P(x, a, x') \left[g(x, a, x') + \beta J^*(x') \right] \\ &= \mathbb{E}[g(x, a, x') + \beta J^*(x')] \\ &\approx \frac{1}{K} \sum_{i=1}^{K} \left(g(x, a, x_i') + \beta J^*(x_i') \right) \end{aligned} \tag{1.18}$$

Note that the argument for optimal Q-factor is a *state-action* pair, which give an estimation of the performance for our action a. Therefore, the optimal policy is defined as

$$\mu^*(x) = \arg \min_{a \in \mathcal{A}(x)} Q^*(x, a) \tag{1.19}$$

Remark 1.4. When Q^* is too big or solving (1.19) is too difficult, we need a low dimensional representation of Q^*.

Bellman equation for Q-factor The definition (1.18) motivates to define a way to compute Q-factor by the Bellman equation. First define

$$(TQ)(x, a) = \sum_{y \in \mathcal{X}} P(x, a, y) \left[g(x, a, y) + \beta \min_{v \in \mathcal{A}(y)} Q(y, v) \right] \tag{1.20}$$

Then Q^* is the unique fixed point to the equation $Q = TQ$.

Remark 1.5. When both P and g are known, the value iteration $Q_{k+1} = TQ_k$ converges to Q^* from any staring Q_0; otherwise we need to sample to approximate P, and therefore the convergence may not be guaranteed.

Also, the policy obtained in each iteration is a deterministic value, and therefore we can introduce a stochastic policy evaluation algorithm.

Q-Learning as a stochastic version

1. First generate long sequence of $\{x_k, a_k, g_k\}$ s.t. the state-action pair (x, a) appears infinitely often.

2. Then the Q-factor of (x_k, a_k) pair is updated:

$$\begin{cases} Q_{k+1}(x_k, a_k) = (1 - \gamma_k) Q_k(x_k, a_k) + \gamma_k \left(g_k + \beta \min_v Q_k(x'_{k+1}, v) \right) \\ Q_{k+1}(x, a) = Q_k(x, a), \text{ if } (x, a) \neq (x_k, a_k) \end{cases} \tag{1.21}$$

3. Note that $g_k + \beta \min_v Q_k(x'_{k+1}, v)$ is a single sample approximation of the expected value $(TQ)(x_k, a_k)$.

The literature [Watkins and Dayan, 1992] gives a convergence result for the Q-learning:

Theorem 1.6. Given the condition that:

1. A sequence where each state-action pair appears *infinitely often*
2. The stage cost is bounded, i.e., $|g(x,a,y)| < G$
3. The learning rate $\gamma_k \in (0,1)$ is such that $\sum_{k=0}^{\infty} \gamma_k = \infty, \sum_{k=0}^{\infty} \gamma_k^2 < \infty$

The Q-learning algorithm converges:

$$\lim_{k\to\infty} Q_k(x,a) = Q^*, \text{ a.s.}$$

Remark 1.6. The theory of reinforcement learning requires ambitious condition, and in practice we can only make the approximation. Therefore, most of algorithms are heuristic.

1.3.3 Policy Iteration using Q-factors

- (Policy Evaluation): Given current policy μ^k, we find the fixed point Q_{μ^k} for the equation

$$Q(x,a) = \sum_{y\in\mathcal{X}} P(x,a,y)[g(x,a,y) + \beta Q(y,\mu^k(y))]$$

- (Policy Improvement): Find $\mu^{k+1}(x)$ based on Q_{μ^k}:

$$\mu^{k+1}(x) = \arg\min_{a\in\mathcal{A}(x)} Q_{\mu^k}(x,a)$$

Here the question turns out that how to make the policy evaluation regarding Q:

SARSA Heuristic Algorithm

1. (Initialization): arbitrary initialize Q^0, choose initial state x_0, initial decision a_0
2. (Simulation): simulate one step starting from x_k, with the decision given by a_k. Observe the next state x_{k+1} and the cost g_k

3. (Evaluation & Improvement):

$$a_{k+1} = \begin{cases} \arg\min_{a \in \mathcal{A}} Q^k(x_{k+1}, a), & \text{w.p. } 1 - \varepsilon \\ \text{other action}, & \text{w.p. } \varepsilon \end{cases}$$

4. (Update):

$$\begin{cases} Q^{k+1}(x_k, a_k) = (1 - \gamma_k) Q^k(x_k, a_k) + \gamma_k \left[g_k + \beta Q^k(x_{k+1}, a_{k+1}) \right], \\ Q^{k+1}(y, v) = Q^k(y, v), \quad \text{for } (y, v) \neq (x_k, a_k) \end{cases}$$

5. $k = k + 1$, move into step 2.

$SARSA$: State $\rightarrow$ Action $\rightarrow$ Reward $\rightarrow$ State $\rightarrow$ Action

2

The Distribution and Installation

2.1 Pre-requisites

You will need a working LaTeX installation. We recomend using pdflatex to process the files. You will also need biber.exe installed. This is distributed as part of the latest versions of LiveTex and MikTex. If you have problems, please let us know.

2.2 The Distribution

The distribution contains 2 folders: `nowfnt` and `nowfnttexmf`.

2.2.1 Folder `nowfnt`

This folder contains the following files using a flat stucture required to compile a FnT issue:

- essence_logo.eps
- essence_logo.pdf
- now_logo.eps
- now_logo.pdf

- nowfnt.cls
- nowfnt-biblatex.sty
- NOWFnT-data.tex

It also contains the following folders:

journaldata A set of data files containing the journal-specific data for each journal. There are three files per journal:
<jrnlcode>-editorialboard.tex
<jrnlcode>-journaldata.tex
<jrnlcode>-seriespage.tex
<jrnlcode> is the code given in Appendix A. You will need these three files to compile your article.

SampleArticle This folder contains this document as an example of an article typeset in our class file. The document is called `FnTarticle.tex`. It also contains this PDF file and the .bib file.

2.2.2 Folder `nowfnttexmf`

This folder contains all the files required in a texmf structure for easy installation.

2.3 Installation

If your LaTeX installation uses a localtexmf folder, you can copy the nowtexmf folder to the localtexmf folder and make it known to your TeX installation. You can now proceed to use the class file as normal.

If you prefer to use the flat files, you will need to copy all the required files each time into the folder in which you are compiling the article. Do not forget to copy the three data files for the specific journal from the folder `journaldata`.

You may need to configure your TeX editor to be able to run the programs. If you have problems installing these files in your own system, please contact us. We use Computer Modern fonts for some of the

journals. You will need to make sure that these fonts are installed. Refer to your system documentation on how to do this.

3

Quick Start

The now-journal class file is designed is such a way that you should be able to use any commands you normally would. However, do **not** modify any class or style files included in our distribution. If you do so, we will reject your files.

The preamble contains a number of commands for use when making the final versions of your manuscript once it has been accepted and you have been instructed by our production team.

3.1 \documentclass

The options to this command enable you to choose the journal for which you producing content and to indicate the use of biber.

```
\documentclass[<jrnlcode>,biber]{nowfnt}.
```

<jrnlcode> is the pre-defined code identifying each journal. See Appendix A for the appropriate <jrnlcode>.

3.2 \issuesetup

These commands are only used in the final published version. Leave these as the default until our production team instructs you to change them.

3.3 \maintitleauthorlist

This is the authors list for the cover page. Use the name, affilliation and email address. Separate each line in the address by \\.

Separate authors by \and. Do not use verbatim or problematic symbols. _ (underscore) in email address should be entered as _. Pay attention to long email addresses.

If your author list is too long to fit on a single page you can use double column. In this case, precede the \maintitleauthorlist command with the following:

```
\booltrue{authortwocolumn}
```

3.4 \author and \affil

These commands are used to typeset the authors and the afilliations on the abstract page of the article and in the bibliographic data.

\author uses an optional number to match the author with the affiliation. The author name is written <surname>, <firstname>.

\affil uses an optional number to match the author with the author name. The content is <affililiation>; <email address>.

3.5 \addbibresource

Use this to identify the name of the bib file to be used.

4

Style Guidelines and LaTeX Conventions

In this section, we outline guidelines for typesetting and using LaTeX that you should follow when preparing your document

4.1 Abstract

Ensure that the abstract is contained within the

`\begin{abstract}`

environment.

4.2 Acknowledgements

Ensure that the acknowledgements are contained within the

`\begin{acknowledgements}`

environment.

4.3 References

now publishers uses two main reference styles. One is numeric and the other is author/year. The style for this is pre-defined in the LaTeX

distribution and must not be altered. The style used for each journal is given in the table in Appendix A. Consult the sample-now.bib file for an example of different reference types.

The References section is generated by placing the following commands at the end of the file.

```
\backmatter
\printbibliography
```

4.4 Citations

Use standard \cite, \citep and \citet commands to generate citations.

Run biber on your file after compiling the article. This will automatically create the correct style and format for the References.

4.4.1 Example citations

This section cites some sample references for your convenience. These are in author/year format and the output is shown in the References at the end of this document.

Example output when using `citet`: **arvolumenumber** is a citation of reference 1 and Bertsekas (1995) is a citation of reference 2.

Example output when using `citep`: (**beditorvolumenumber**) is a citation of reference 3 and (**inproceedings**) is a citation of reference 4.

4.5 Preface and Other Special Chapters

If you want to include a preface, it should be defined as follows:

```
\chapter*{Preface}
\markboth{\sffamily\slshape Preface}
  {\sffamily\slshape Preface}
```

This ensures that the preface appears correctly in the running headings.

You can follow a similar procedure if you want to include additional unnumbered chapters (*e.g.*, a chapter on notation used in the paper), though all such chapters should precede Chapter 1.

Unnumbered chapters should not include numbered sections. If you want to break your preface into sections, use the starred versions of `section`, `subsection`, *etc.*

4.6 Long Chapter and Section Names

If you have a very long chapter or section name, it may not appear nicely in the table of contents, running heading, document body, or some subset of these. It is possible to have different text appear in all three places if needed using the following code:

```
\chapter[Table of Contents Name]{Body Text Name}
\chaptermark{Running Heading Name}
```

Sections can be handled similarly using the `sectionmark` command instead of `chaptermark`.

For example, the full name should always appear in the table of contents, but may need a manual line break to look good. For the running heading, an abbreviated version of the title should be provided. The appearance of the long title in the body may look fine with LaTeX's default line breaking method or may need a manual line break somewhere, possibly in a different place from the contents listing.

Long titles for the article itself should be left as is, with no manual line breaks introduced. The article title is used automatically in a number of different places by the class file and manual line breaks will interfere with the output. If you have questions about how the title appears in the front matter, please contact us.

4.7 Internet Addresses

The class file includes the `url` package, so you should wrap email and web addresses with `\url{}`. This will also make these links clickable in the PDF.

5

Compiling Your FnT Article

During the first run using the class file, a number of new files will be created that are used to create the book and ebook versions during the final production stage. You can ignore these until preparing the final versions as described in Section 5.3. A complete list of the files produced are given in Appendix B.

5.1 Compiling Your Article Prior to Submission

To compile an article prior to submission proceed as follows:

Step 1: Compile the LaTeX file using pdflatex.

Step 2: Run biber on your file.

Step 3: Compile again using pdfLaTeX. Repeat this step.

Step 4: Inspect the PDF for bad typesetting. The output PDF should be similar to FnTarticle.pdf. Work from the first page when making adjustments to resolve bad line breaks and bad page breaks. Rerun pdfLaTeX on the file to check the output after each change.

5.2 Preparing the Final Versions

If you choose the option to compile the final versions of your PDF for publication, you will receive a set of data from our production team upon final acceptance. With the exception of "lastpage", enter the data into the `\issuesetup` command in the preamble.

lastpage= This is the last page number in the sequential numbering of the journal volume. You will need to enter this once you have compiled the article once (see below).

5.3 Compiling The Final Versions

The final versions should be created once you received all the bibliographic data from our Production Team and you've entered it into the preamble. You will be creating a final online journal version pdf; a printed book version pdf; and an ebook version pdf.

Step 1: Compile the LaTeX file using pdfLaTeX.ArAuthor *et al.*, 2014

Step 2: Run biber on your file.

Step 3: Compile again using pdfLaTeX. Repeat this step.

Step 4: Inspect the PDF for bad typesetting. The output PDF should be similar to FnTarticle.pdf. Work from the first page when making adjustments to resolve bad line breaks and bad page breaks. Rerun pdfLaTeX on the file to check the output after each change.

Step 5: When you are happy with the output make a note of the last page number and enter this in `\issuesetup`.

Step 6: Compile the article again.

Step 7: Open the file <YourFilename>-nowbook.tex. This will generate the printed book version pdf.

Step 8: Compile the LaTeX file using pdflatex.

Step 9: Run biber on your file.

Step 10: Compile again using pdfLaTeX. Repeat this step.

Step 11: Repeat steps 7-10 on the file: <YourFilename>-nowebook.tex. This will generate the ebook version pdf.

Step 12: Repeat steps 7-10 on the file: <YourFilename>-nowplain.tex. This will generate a plain version pdf of your article. If you intend to post your article in an online repository, please use this version.

5.4 Wednesday

Proposition 5.1. G is hamiltionian implies that for any nonempty $S \subseteq V$, $G - S$ has at most $|S|$ components.

Theorem 5.1. G is Hamiltionian if for any vetices v, w such that $(v, w) \notin E$,

$$\deg(v) + \deg(w) \geq n$$

Knapsacle Problem

$$\begin{array}{ll} \max & \sum_{i=1}^{n} v_i u_i \\ \text{such that} & \sum_{i=1}^{n} w_i u_i \leq W \\ & u_i \in \{0,1\},\ i = 1, \ldots, n \end{array}$$

This is the binary integer programming problem, which is NP-complete. We cannot find exact solution to this problem. There are 2^n possible solutions, which requires exponential time.

We can find some "pseudo"-polynomial algorithm. Assumption:

1. W is an integer.

State: remaining capacity, define as X_k.
Stage: iterm $1, \ldots, n$.

$$J_N(x_N) = \begin{cases} v_N & \text{if } N \leq X_N \\ 0, & \text{if } w_N > X_N \end{cases}$$

$$J_N(x_{N-1}) = \begin{cases} 0 + J_N(X_N), & \text{if } w_{N-1} > X_{N-1}, \text{Here } X_N = \\ \max\{v_{N-1} + J_N(x_N), 0 + J_N(x_N)\}, & \text{if } w_{N-1} \leq X_{N-1} \end{cases}$$

Here why the DP has "pesduo"-polynomial time, and the w_i should be integer?

Let's list states $\{x_1, \ldots, x_N\}$. For each stage k, there would be $W+1$.

Each iteration at most 2 computation, and therefore we face $2(W+1) \cdot N$

5.4.1 Label Correcting Methods

Shortest Path. Back up some information.

The label refers to the intermediate information.

A^*-algorithm; Bellman-Ford Algorithm

Benchmark: MILP, CPLEX, CVX.

5.4.2 DP problems with perfect state information

Linear Quadratic System Let $x_k \in \mathbb{R}^1$ be the state; $u_k \in \mathbb{R}^n$ be the control; $\omega_k \in \mathbb{R}^n$ be the disturbance.

System Dynamics.

$$x_{k+1} = A_k x_k + B_k u_k + \omega_k$$

Stage cost:

$$x_k' Q_k x_k + u_k' R_k u_k$$

Here Q_k is required to be positive-semi-definite symmetric matrix; R_k to be positive-definite symmetric matrix.

$$J_{n-1}(x_{n-1}) = \min_{u_{n-1}} \mathbb{E}_\omega \left\{ x_{n-1}' Q_{n-1} x_{n-1} + u_{n-1}' R_{n-1} u_{n-1} + J_n(x_n) \right\}$$

where

$$J_n(x_n) = (A_{n-1} x_{n-1} + B u_{n-1} + \omega)' Q_n (A_{n-1} x_{n-1} + B u_{n-1} + \omega)$$

5.4.3 Linear Quadratic

Dynamics:

$$x_{k+1} = A_k x_k + B_k u_k + \omega_k$$

Stage cost:

$$x_k' Q_k x_k + u_k' R_k u_k, \ Q_k \succeq 0, R_k \succ 0$$

Cost to go function:

$$J_k(x_k) = \min_{u_k} \mathbb{E}_{\omega_k}[x_k' Q_k x_k + u_k' R_k u_k + J_{k+1}(x_{k+1})]$$

The final cost is

$$J_N(x_N) = x_N' Q_N x_N$$

At $N-1$ stage

$$x_{N-1}' Q_{N-1} x_N + u_{N-1}' R_{N-1} u_{N-1} + \mathbb{E}[(A_{N-1}x_{N-1} + B_{N-1}u_{N-1} + \omega_{N-1})' Q_N (A_{N-1}x_{N-1} + B_{N-1}u_{N-1} + \omega_{N-1})]$$

Optimization model:

$$\frac{\partial}{\partial u}(u'Hu + 2r'u + c) = 2Hu + 2r \implies Hu = -r \implies u^* = -H^{-1}r$$

For $N-1$ stage, we have

$$\begin{aligned} H &= R_{N-1} + B_{N-1}' Q_N B_{N-1} \\ r' &= x_{N-1}' A_{N-1}' Q_N B_N \\ c &= x_{N-1}'(Q_{N-1} + A_{N-1}' Q_N A_{N-1}) x_{N-1} + \mathbb{E}_\omega(\omega_{N-1}' Q_N \omega_{N-1}) \end{aligned}$$

Therefore,

$$u_{N-1}^* = -(R_{N-1} + B_{N-1}' Q_N B_{N-1})^{-1} B_N Q_N A_{N-1} x_{N-1},$$

which is linaer in terms of x_{N-1}, which is so called linear controller.

Therefore,

$$J_{N-1}(x_{N-1}) = x_{N-1}' K_{N-1} x_{N-1} + \mathbb{E}(\omega_{N-1}' Q_N \omega_{N-1})$$

The linear controller will be tested during mid-term or final.

$$J_0(x_0) = x_0' K_0 x_0 + \sum_{k=0}^{N-1} \mathbb{E}_\omega\{\omega_k' K_{k+1} \omega_k\}$$

$$K_{k-1} = f(K_k) \implies K_{k-1} = K_k, \quad \text{for large } k.$$

Riccati Equation for stationary ststem, i.e., $A_k = A, B_k = B, Q_k = Q, R_k = R$.

$$K = A'(K - KB(B'KB + R)^{-1}B'K)A + Q$$

Therefore as time goes long enough, the cost to go function $J_k(x_k)$ will be a constant plus over stages. such a constant K is called the *optimal stationary controller*.

Stability. For $u^* = Lx$, we imply

$$x_{k+1} = Ax_k + Bu_k + \omega_k = (A + BL)x_k + \omega_k$$

Care about $\lim_{k\to\infty}(A + BL)^k = 0$.

Here

$$L = -(B'KB + R)^{-1}B'KA$$

It suffices to solve

$$P_{k+1} = A^2\left(P_k - \frac{B^2P_k^2}{B^2P_k + R}\right) + Q$$

Then consider the case that A_k, B_k are all random matrices, i.e., independent. $Q_k \succeq 0, R_k \succ 0$.

$$L_k = -\left(R_k + \mathbb{E}(B_k'K_{k+1}B_k)\right)^{-1}\mathbb{E}(B_j'K_{k+1}A_k)$$

Note that

$$P_\infty = \frac{\mathbb{E}A^2RP}{R + EB^2P} + \frac{\mathbb{E}A^2\mathbb{E}B^2 - (\mathbb{E}A)^2(\mathbb{E}B)^2}{R + \mathbb{E}B^2P}$$

Certainty Equivalence
State: x_k, inventory level
Control: $u_k \geq 0$, number of orders placed
Disturbance: ω_k: damand
Dynamics:

$$x_{k+1} = (x_k + u_k - \omega_k)$$

Stage cost:

- Ordering cost: $c \cdot u_k$

- Maintaining cost: full backlog.$\mathbb{E}_{\omega_k}(r(x_k + u_k - \omega_k))$.

  where $r(z) = hz^+ + pz^-$ is a convex function.

  We imply the maintaining cost is $H(x_k + u_k)$, which is convex.

$$J_k(x_k) = \min_{u_k \geq 0} \{cu_k + H(x_k + u_k) + \mathbb{E}[J_{k+1}(x_k + u_k - \omega_k)]\}$$

Define $Y = x_k + u_k$, and therefore

$$G(Y) = cY + H(Y) + \mathbb{E}[J(Y - \omega_k)]$$

Therefore, we can always set $Y = x_k + u_k = S_k$, where S_k minimizes $G(Y)$.

Acknowledgements

The authors are grateful to Ulrike Fischer, who designed the style files, and Neal Parikh, who laid the groundwork for these style files.

Appendices

A

Journal Codes

The table below shows the journal codes to be used in `\documentclass`. For Example: `\documentclass[ACC,biber]{nowfnt}`

Journal	<jrnlcode>	**R**ef. Style
Annals of Corporate Governance	ACG	Author/Year
Annals of Science and Technology Policy	ASTP	Author/Year
FnT Accounting	ACC	Author/Year
FnT Comm. and Information Theory	CIT	Numeric
FnT Databases	DBS	Author/Year
FnT Econometrics	ECO	Author/Year
FnT Electronic Design Automation	EDA	Author/Year
FnT Electric Energy Systems	EES	Author/Year
FnT Entrepreneurship	ENT	Author/Year
FnT Finance	FIN	Author/Year
FnT Human-Computer Interaction	HCI	Author/Year
FnT Information Retrieval	INR	Author/Year
FnT Information Systems	ISY	Author/Year
FnT Machine Learning	MAL	Author/Year
FnT Management	MGT	Author/Year
FnT Marketing	MKT	Author/Year
FnT Networking	NET	Numeric

Continues

Journal	<jrnlcode>	**R**ef. Style
FnT Optimization	OPT	Numeric
FnT Programming Languages	PGL	Author/Year
FnT Robotics	ROB	Author/Year
FnT Privacy and Security	SEC	Author/Year
FnT Signal Processing	SIG	Numeric
FnT Systems and Control	SYS	Author/Year
FnT Theoretical Computer Science	TCS	Numeric
FnT Technology, Information and OM	TOM	Author/Year
FnT Web Science	WEB	Author/Year

B

Files Produced During Compilation

The files that are created during compilation are listed below. The additional *.tex files are used during the final production process only. See Section 5.3.

```
<YourFilename>-nowbook.tex
<YourFilename>-nowchapter.tex
<YourFilename>-nowebook.tex
<YourFilename>-nowechapter.tex
<YourFilename>-nowsample.tex
<YourFilename>-nowplain.tex
<YourFilename>.aux
<YourFilename>.bbl
<YourFilename>.bcf
<YourFilename>.blg
<YourFilename>.log
<YourFilename>.out
<YourFilename>.pdf
<YourFilename>.run.xml
<YourFilename>.synctex.gz
<YourFilename>.tex
<YourFilename>.toc
```

References

ArAuthor, A., B. Author, and C. Author (2014). "An article on something interesting (volume only)". *Journal of interesting Things*. 6: 1–122. I have something special to say about this publication. ISSN: 0899-8256. DOI: 10.5161/202.00000013. URL: http://www.nowpublishers.com/ (accessed on 07/10/2014).

Bertsekas, D. (1995). "Dynamic Programming and Optimal Control". In: vol. 1.

Sutton, R. S. (1988). "Learning to predict by the methods of temporal differences". *Machine Learning*. 3: 9–44.

Watkins, C. J. C. H. and P. Dayan (1992). "Q-learning". In: *Machine Learning*. 279–292.