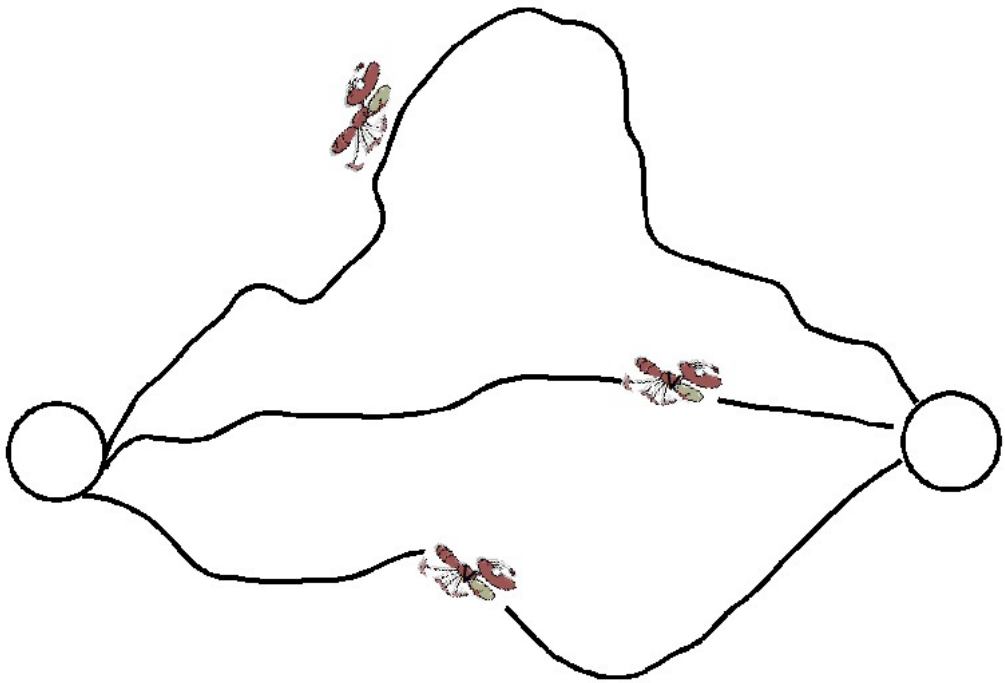


Vivek S. Borkar
TIFR, Mumbai.

NAIVE ANTS



FORAGING BEHAVIOR OF ANTS:

1. Ants explore several alternative paths to food.
2. Ants deposit pheromone while they traverse the path.
3. On finding food, they return to base.
4. Shorter paths \implies more ants return early
 \implies more pheromone on the path
5. Ants prefer paths of higher pheromone (i.e., assign them a higher probability).
6. \implies Ants concentrate on the shortest path.

‘POPULATION ALGORITHM’

Motivation: Traditional optimization algorithms: deterministic, use local information \implies get local optima \implies for global optimum, give up locality (e.g., in multi-start), or determinism (\implies go stochastic, e.g., in simulated annealing)

Population algorithms are non-local (use a whole population of concurrently running, *interacting* algorithms) *and* stochastic,

Use stochasticity to explore and some reinforcement mechanism to accentuate desired behaviour. (Examples: ACO, PSO, genetic algorithms)

IMPORTANT OBSERVATION ABOUT ACO:

Initial bias leads to convergence to the desired behavior.

Question: Is initial bias alone good enough?

Answer: Not quite!

Aside: Contrast with ‘locking in’ phenomenon of ‘increasing returns’ economics: In ACO, initial conditions are a blessing rather than a hindrance!

Conventional schemes based on the ‘**Pheromone Update Rule**’: For i -th path,

$$\tau_i(n + 1) = (1 - \rho)\tau_i(n) + \rho\zeta_i(n + 1).$$

Here:

- $\rho \in (0, 1)$ a ‘forgetting factor’ or ‘pheromone evaporation factor’ (necessary for avoiding getting trapped into undesirable behavior).
- $\zeta_i(n + 1) =$ the number of ants joining path i , with,

$$\begin{aligned}\zeta_i(n + 1) &= R \text{ with prob. } \frac{\tau_i(n)}{\sum_j \tau_j(n)}, \\ &= 0 \quad \text{otherwise.}\end{aligned}$$

Empirically, the pure pheromone update does not do too well. In particular, it needs a very large number of ants to see good results.

Quickfix: ‘*Elitist*’ strategy: Keep track of the best so far.

Fact: converges to the desired behavior (trivially!)

A better approach:

1. Change pheromone update to:

$$\tau_i(n+1) = (1 - \rho)\tau_i(n) + \rho R Q_i(n),$$

where $Q_i(n)$ = the number of ants returning on path i at time n .

2. Introduce ‘*Agent Learning Parameter*’ $\{X_i(n)\}$: For $a(n) > 0$, $\sum_n a(n) = \infty$, $\sum_n a(n)^2 < \infty$,

$$X_i(n+1) = X_i(n) + a(n)X_i(n)\tau_i(n+1).$$

3. Change the (conditional) prob. that $\zeta_i(n+1) = R$ to:

$$P_i(n) \stackrel{def}{=} \frac{X_i(n)}{\sum_j X_j(n)}.$$

Can show:

$$P_i(n+1) - P_i(n) =$$

$$a(n)P_i(n) \left[\tau_i(n+1) - \sum_j P_j(n)\tau_j(n+1) \right] + o(a(n)).$$

Limiting o.d.e. is the *replicator dynamics*:

$$\dot{p}_i(t) = p_i(t) \left[f_i(p(t)) - \sum_j p_j(t) f_j(p(t)) \right],$$

where $f_i(p) \approx$ the ‘stationary average of $\tau_i(n)$ when $p_n \equiv p \forall n$ ’.

Algorithm tracks the o.d.e. asymptotically with probability 1.

WHY?

See:

V. S. Borkar, '*Stochastic Approximation: A Dynamical Systems Viewpoint*', Hindustan Book Agency and Cambridge Uni. Press, 2008.

(Unofficially subtitled –
'*Everything you never cared to know about stochastic approximation and did not bother to ask*')

Need: $f_i(p) \uparrow$ as $p_i \uparrow$.

Claim: $[1, 0, \dots, 0]$ is an asymptotically stable equilibrium for the ‘*replicator dynamics*’

$$\dot{p}_i(t) = p_i(t)[f_i(p(t)) - \sum_j p_j(t) f_j(p(t))],$$

with the domain of attraction

$$\mathcal{D}_1 = \{[p_1, \dots, p_d] : p_1 > p_j, j \neq 1\}.$$

Symmetric results hold for other corners of the probability simplex \implies initial bias gets reinforced. (Remember ‘MAX-NET’.)

Can also show: If $P(n) \in \mathcal{D}_1$

$$P(P_k \xrightarrow{n \leq k \uparrow \infty} [1, 0, \dots, 0]) \geq 1 - K \sum_{k \geq n} a(k)^2,$$

where K depends on $n, \|P(n)\|, \|\tau(n)\|$

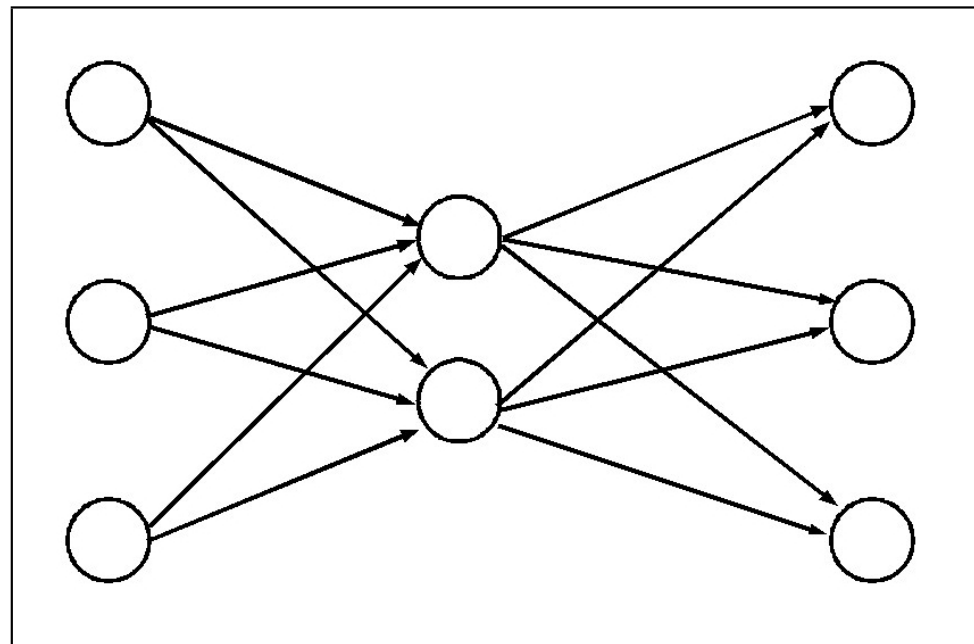
Empirically, does better than the naive scheme, does not require keeping track of the ‘best so far’.

Choice of ρ not very critical. Choice of $a \approx$ tradeoff between speed and fluctuations. Choice of N : **Critical**, significant improvement as $N \uparrow \infty$. (Precise estimates possible using ‘large deviations’ theory)

Extensions: Multistage shortest path

Ants do '*Dynamic Programming*'!

Key idea: Same arguments as before for the last stage, thereafter do backward induction.



Extensions: TSP

Ants do a 'self-avoiding' walk by keeping a memory of nodes visited so far.

Can map to a multistage shortest path for analysis purposes. Convergence does not follow as before, because the same edge can collect pheromone from different paths.

Normalization by path length: improves performance.

CONGESTION-SENSITIVE ANTS

The model

- Users arrive at a ‘source node’, choose one of M routes.
- $X_i(n) \stackrel{def}{=} \#$ of users already on route i at time n ;
- $Y_i(n) = U_i(X_i(n))$, utility perceived by a newly arriving customer at time n for route i .
- $\xi_i(n) (\in [0, 1]) \stackrel{def}{=} \text{i.i.d. random fraction of users on route } i \text{ completing their journey at time } n,$
 $E[\xi_i(n)] = \mu > 0.$
- Arrival process $\{\zeta(n)\}$ i.i.d. with mean λ .
- $\gamma_i(n) \in [0, 1] =$ the fraction of new arrivals at time n that join the i -th route.

Assume:

1. $U_i(\cdot) \geq 0$ differentiable, monotone decreasing.
2. U_i 's have identical growth properties, i.e.,

$$U_i(x) = \Theta(U_j(x)) \text{ as } \|x\| \rightarrow \infty, \quad i \neq j. \quad (1)$$

A newly arriving user joins route i with probability

$$\frac{U_i(X_i(n))^\alpha}{\sum_j U_j(X_j(n))^\alpha}, \quad \alpha \geq 1. \quad (2)$$

Assume: μ small, $\lambda < \mu M$ (ensures stability).

Want: $\mu \downarrow 0$ limit with $\frac{\lambda}{\mu}$ kept constant ('*small user*').

Dynamics of $\{X_i(n)\}$:

$$\begin{aligned} X_i(n+1) &= X_i(n) + \gamma_i(n+1)\zeta(n+1) - \xi_i(n+1)X_i(n) \\ &= X_i(n) + \mu \left(\frac{\lambda}{\mu} \frac{U_i(X_i(n))^\alpha}{\sum_j U_j(X_j(n))^\alpha} - X_i(n) \right) \\ &\quad + M_i(n+1), \end{aligned}$$

where,

$$\begin{aligned} M_i(n+1) &\stackrel{def}{=} (\gamma_i(n+1)\zeta(n+1) - \xi_i(n+1)X_i(n)) \\ &\quad - \left(\lambda \frac{U_i(X_i(n))^\alpha}{\sum_j U_j(X_j(n))^\alpha} - \mu X_i(n) \right), n \geq 0. \end{aligned}$$

(a martingale difference sequence)

This dynamics can be viewed as a constant stepsize ($= \mu$) stochastic approximation algorithm.

\implies it follows the asymptotic behavior of the o.d.e.

$$\begin{aligned}\dot{x}_i(t) &= h_i(x(t)) - x_i(t) \\ &= \left(\frac{\lambda}{\mu}\right) \frac{U_i(x_i(t))^\alpha}{\sum_j U_j(x_j(t))^\alpha} - x_i(t), \quad 1 \leq i \leq M, t \geq 0.\end{aligned}$$

Let $h(\cdot) = [h_1(\cdot), \dots, h_M(\cdot)]^T$.

$$\frac{\partial h_i(x)}{\partial x_j} = \frac{\lambda}{\mu} \left[- \frac{\alpha U_i(x_i)^\alpha U'_j(x_j) U_j(x_j)^{\alpha-1}}{(\sum_k U_k(x_k)^\alpha)^2} \right] > 0, \quad i \neq j.$$

\implies *cooperative* o.d.e. Since

$$x_i(t) = e^{-t} x_i(0) + \int_0^t e^{-(t-s)} h_i(x(s)) ds, \quad 1 \leq i \leq M, \quad t \geq 0,$$

and $|h_i(\cdot)|$ are bounded, $x(\cdot)$ remains bounded.

\implies by a theorem of Hirsch, we have:

For initial conditions belonging to an open dense set, $x(\cdot)$ converges to the set $H_\alpha \stackrel{\text{def}}{=} \{x : h(x) = x\}$.

By a theorem of Borkar and Meyn,
 $\sup_n E[\|X(n)\|^2] < \infty \implies \{X(n)\}$ is a Markov chain with
at least one invariant distribution.

Assumption: (*) There exists an open dense $D \subset \mathcal{R}^M$
such that $\forall x(0) \in D, x(t) \rightarrow H_\alpha$; and *any* invariant
distribution of $\{X(n)\}$ assigns zero probability to D^c .

Intuitively, D^c should comprise of all points excluded in
the Hirsch result.

For $\epsilon > 0$, let $H_\alpha^\epsilon \stackrel{def}{=} \{x : \inf_{y \in H_\alpha} \|x - y\| < \epsilon\}$.

We can use standard proof techniques for constant step-size stochastic approximation to conclude:

Theorem For any $\epsilon, \delta \in (0, 1)$,

$$\limsup_{n \uparrow \infty} P(X(n) \notin H_\alpha^\epsilon) \leq \delta + O\left(\frac{\mu}{\epsilon}\right).$$

That is, if μ is ‘small’, then $\{X(n)\}$ asymptotically concentrate near H_α with high probability.

Wardrop equilibria

Let

$$H_\infty \stackrel{\text{def}}{=} \{x = [x_1, \dots, x_M] : U_i(x_i) \neq \max_j U_j(x_j) \Rightarrow x_i = 0 \\ \text{and } U_k(x_k), U_\ell(x_\ell) = \max_j U_j(x_j) \Rightarrow x_k = x_\ell\}.$$

Theorem As $\alpha \uparrow \infty$, $H_\alpha \rightarrow H_\infty$ in the sense that, if $x_\alpha \in H_\alpha \forall \alpha$ and $x_\alpha \rightarrow x_\infty$ along a subsequence as $\alpha \uparrow \infty$, then $x_\infty \in H_\infty$.

Proof: Let

$$S(x) \stackrel{def}{=} \{i : U_i(x_i) = \max_j U_j(x_j)\}, x = [x_1, \dots, x_M].$$

As $\alpha \uparrow \infty$,

$$\begin{aligned} \frac{U_i(x_i)^\alpha}{\sum_j U_j(x_j)^\alpha} &\rightarrow \frac{1}{|S(x)|}, \quad i \in S(x), \\ &\rightarrow 0, \quad i \notin S(x). \end{aligned}$$

The claim follows. □

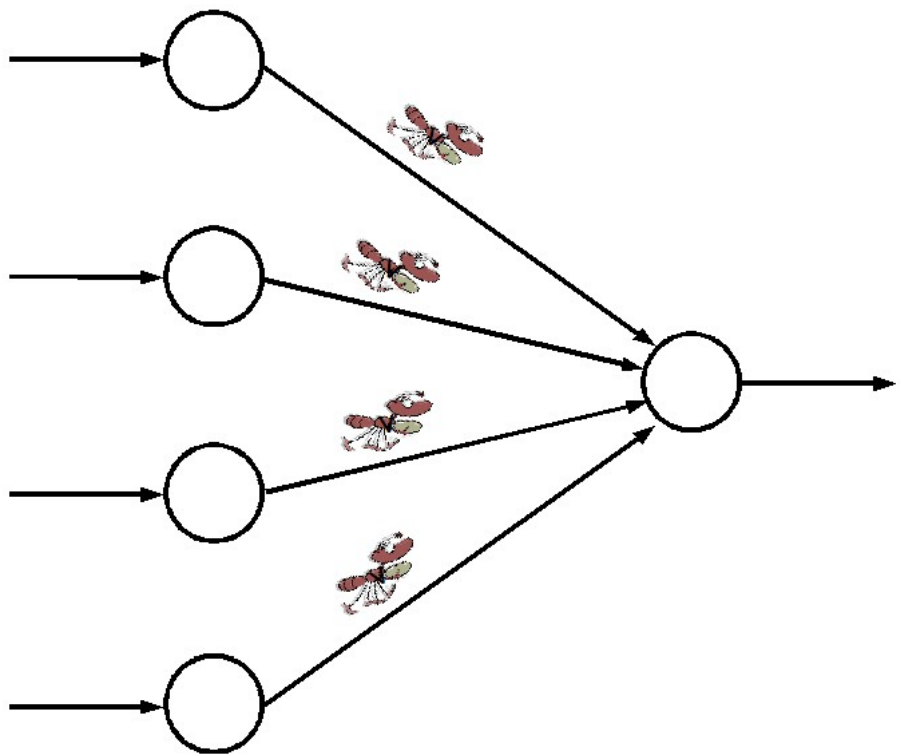
The set corresponds to equal allocation of flow to the utility maximizing routes, with zero allocation to the rest. This corresponds to the notion of **Wardrop equilibria** from transportation theory.

Multi-stage version of congestion dependent transport possible!

Use estimated round trip delays as measure of disutility, it is additive over stages. Argue as for naive ants.

Needs *separation of time scales* for convergence.

COMPETING ANTS



$A_k(t) := \#$ of arrivals of type k in slot t (i.i.d.)

$U_k(t) := \#$ of type k served in slot t .

$$R_k(t) = \sum_{s=0}^t U_k(s).$$

$X_k(t) :=$ the queue length of type k at the beginning of slot t . Then,

$$X_k(t+1) = X_k(t) - U_k(t) + A_k(t)$$

$$U_k(t) \leq X_k(t)$$

Constraint: $\limsup_{t \rightarrow \infty} \frac{1}{t} E[\sum_{k=1}^K R_k(t)] \leq R.$

AIM: Minimize $\limsup_{t \rightarrow \infty} \frac{1}{t} E[\sum_{s=0}^{t-1} \sum_{k=1}^K X_k(s)]$

‘Constrained MDP’: LP formulation (primal)

$$\begin{aligned} \min \sum_{(x_k, u_k)} \sum_{k=1}^K c_k(x_k, u_k) \mu^{(k)}(x_k, u_k) & \quad \text{subject to} \\ \sum_{(y_k, u_k)} \mu^{(k)}(y_k, u_k) p^{(k)}(x_k | y_k, u_k) & = \sum_{u_k} \mu^{(k)}(x_k, u_k), \\ - \sum_{k=1}^K \sum_{(x_k, u_k)} u_k \mu^{(k)}(x_k, u_k) & \geq -R \\ \sum_{(x_k, u_k)} \mu^{(k)}(x_k, u_k) & = 1, \\ \mu^{(k)}(x_k, u_k) & \geq 0 \quad \forall (x_k, u_k). \end{aligned}$$

Dual LP:

$$\begin{aligned} \max \left(\sum_{k=1}^K \beta_k - \lambda R \right) \quad & \text{subject to} \\ \beta_k + V^{(k)}(x_k) - \lambda u_k & \leq c_k(x_k, u_k) + \\ & \sum_{y_k} p^{(k)}(y_k | (x_k, u_k)) V^{(k)}(y_k), \\ \lambda & \geq 0 \end{aligned}$$

Separates into individual LPs if λ known! (Dynamic counterpart of Kelly's decomposition)

Can calculate λ by

$$\lambda_{n+1} = \Gamma\left(\lambda_n + b_n \left(\sum_{k=1}^K u_k^*(\lambda_n, X_k(n)) - R\right)\right),$$

where Γ is a projection operator.

$\lambda_n \rightarrow$ the optimal λ with probability 1 \implies the scheme works.

REFERENCES

1. (V. S. Borkar; D. J. Das) A novel ACO algorithm for optimization via reinforcement and initial bias, *Swarm Intelligence*, to appear.
2. Cooperative dynamics and Wardrop equilibria, *Systems and Control Letters*, to appear.
3. (V. S. Borkar; J. Kuri) Optimal distributed uplink channel allocation: a constrained MDP formulation, submitted to ‘*Annals of Dynamic Games*’.
4. (V. S. Borkar; J. Kuri) Wardrop routing in multi-stage networks, in preparation.

Ladybug: This tastes just like crap.

Beetle: Really? Let me try some.

Hey, it *is* crap. Not bad!