Optimization via Simulation.

Want to minimize $\alpha(\theta)$.

Assuming unimodality and differentiability, want to solve:

$$\alpha'(\theta) \;=\; 0.$$

Example: Min $C(\theta) + w(\theta)$

where $w(\theta)$ is expected average waiting time in a queue,

C is available analytically, and θ is server speed.

Want to solve $C'(\theta) + w'(\theta) = 0$.

If the system is regenerative and $w(\theta) = u(\theta)/\ell(\theta)$,

then $w' \;=\; (\ell u' - u\ell')/\ell^2$ and so we want to solve:

$$\ell^2(\theta)[C'(\theta) + w'(\theta)] \;=\; \ell^2(\theta)C'(\theta) + \ell(\theta)u'(\theta) - u(\theta)\ell'(\theta) \;=\; 0.$$

Remarks:

Optimization in presence of noise: more difficult than deterministic.

Even just estimating the function is difficult: MSE $= O(1/n)$.

Seek a θ that gives improvement; not necessarily the optimum.

Response Surface Methodology (RSM)

Choose initial θ_1, set $n := 1$, and repeat:

— Build a metamodel of $\alpha(\cdot)$ around $\theta_n = (\theta_n^{(1)}, \ldots, \theta_n^{(d)})$.

Usually linear metamodel:

$$\alpha(\theta) = \beta_{n0} + \beta_n^T \theta = \beta_{n0} + \sum_{k=1}^{d} \beta_{nk} \theta^{(k)}.$$

Perform simulations at m_n design points and estimate the β_{nk}'s by least-squares regression.

— Move:

$$\theta_{n+1} := \theta_n - \gamma_n \beta_n.$$

Can perform line search with additional simul. to determine γ_n.

— $n := n + 1$.

Remarks:

Can also fit a quadratic model (e.g., last iteration).

Generally very expensive to build a metamodel for large d.

Gradient surface method (Ho et al. 1992).

Stochastic Counterpart Approach

[See Rubinstein 1991; Rubinstein and Shapiro, 1993].

Idea is to estimate the entire function (and/or derivative) all over Θ from each simulation run, using a likelihood ratio.

Then, minimize the sample function.

SC Algorithm:

Simulate using P to obtain $H(\cdot,\omega)$ all over Θ.

Repeat that N times, to get $H_1(\cdot,\omega_1),\ldots,H_N(\cdot,\omega_N)$. Let

$$\bar{H}_N(\cdot) \;=\; \frac{1}{N}\sum_{i=1}^{N} H_i(\cdot,\omega_i).$$

Compute the minimizer $\hat{\theta}_N$ of $\bar{H}_N(\cdot)$.

Remarks:

1. One may replace H by H' and solve $\bar{H}'_N(\theta)=0$.

   If $\bar{H}_N$ is convex, this is equivalent.

2. Typically, $E\left[\arg\min_\theta H(\cdot,\omega)\right] \;\neq\; \arg\min_\theta E[H(\cdot,\omega)] \;\stackrel{\text{def}}{=}\; \theta_*$.

   So, $\hat{\theta}_N$ is biased for finite N. However:

Theorem (Rubinstein and Shapiro 1993).

Under fairly general conditions, one has $\hat{\theta}_N \to \theta_*$ a.s., and

$$\sqrt{N}(\hat{\theta}_N-\theta_*) \;\Rightarrow\; N(0,\gamma_*\sigma_*^2\gamma_*').$$

where $\gamma_*=[\alpha''(\theta_*)]^{-1}$ and $\sigma_*^2=\mathrm{Var}_{\theta_0}[H'(\theta_*,\omega)]$.

Main practical difficulties:

1. Choice of P.
2. Storing and retrieving the info. required for solving $\bar{H}'_N(\theta) = 0$.
3. If N not large enough, $\bar{H}'_N(\cdot)$ may be badly behaved.
4. Variance and bias of function or gradient estimator.
5. Getting aware that the variance is high for some θ's ....
6. More difficult to use "on-line" than SA.

Suggestion: use at each step of an iterative procedure, as for RSM.

Stochastic Approximation

Iterative algorithm somewhat similar to that of RSM.

Algorithm:

Choose initial θ_1;

FOR $n := 1$ TO N DO

Obtain Y_n, an estimate of $\alpha'(\theta_n)$;

$\theta_{n+1} := \pi_\Theta(\theta_n - \gamma_n Y_n)$;

Here,

π_Θ : projection on Θ.

γ_n : sequence of gains (fixed); [e.g., $\gamma_n = \gamma_0 n^{-1}$];

$Y_n = \alpha'(\theta_n) + B_n + \epsilon_n$; $E[\epsilon_n \mid \theta_n, s_n] = 0$.

Under appropriate conditions, $\theta_n \to \theta_*$ a.s.

E.g., if $\gamma_n = \gamma_0 n^{-1}$, $B_n \stackrel{\text{a.s.}}{\to} 0$, $\sum_{n=1}^{\infty} E[\epsilon_n^2 \mid \theta_1, s_1] n^{-2} < \infty$ a.s.,

$\alpha(\theta)$ strictly convex in Θ, and usual constraints qualifications.

There are also other (weaker) sets of sufficient conditions.

If $\sigma_n^2 = \mathrm{Var}[\epsilon_n \mid \theta_n, s_n] \stackrel{\text{a.s.}}{\to} \sigma^2$, $B_n = 0$, and $\gamma_0 = \gamma_* = [\alpha''(\theta_*)]^{-1}$,

then $\sqrt{n}(\theta_n - \theta_*) \Rightarrow N(0, \gamma_* \sigma^2 \gamma_*')$.

Some practical difficulties:

— Choice of γ_0 (or of sequence γ_n). Nonlinear rescaling ?

— Projection (in some cases).

— Sometimes, variance and bias of gradient estimator.

More on convergence rates.

(Yin and L'Ecuyer 1994).

Let $\gamma_n = \gamma_0 n^{-\gamma}$, $B_n = O(n^{-\beta})$, $V_n = E|Y_n - E_n Y_n|^2 = O(n^{-\delta})$,

$\beta > 0$, $\gamma + \beta > 1$, $2\gamma + \delta > 1$,

and $\gamma_0 \alpha'(\theta_*) > \theta_* \max(2\beta,\ 1+\delta)$ if $\gamma = 1$.

Then $E[(\theta_n - \theta_*)^2] = O(n^{-\kappa})$ where $\kappa = \min(2\beta,\ \gamma+\delta)$.

Suppose that to compute Y_n, we use a computer budget $t_n = O(n^p)$.

$C_n = O(n^{p+1})$ = total budget for first n SA iterations.

Want to know: smallest η such that $E[(\theta_n - \theta_*)^2] = O(C_n^{-\eta})$.

Finite-horizon model: Bias due to finite differences.

Truncated-horizon model: Bias and var. also depend on horizon t_n.

Suppose $B_n = O(t_n^{-b})$ and $V_n = O(t_n^{-d})$.

Then, $\beta = bp$, $\delta = dp$, $\eta = \min(2bp,\ \gamma + dp)/(p+1)$.

model	method	ν	γ	β	δ	p	κ	η
Fin.-h.	FDf	$1/4$	1	$1/4$	$-1/2$	0	$1/2$	$1/2$
Fin.-h.	FDc	$1/6$	1	$1/3$	$-1/3$	0	$2/3$	$2/3$
Fin.-h.	FDCf-I	$1/3$	1	$1/3$	$-1/3$	0	$2/3$	$2/3$
Fin.-h.	FDCc-I	$1/5$	1	$2/5$	$-1/5$	0	$4/5$	$4/5$
Fin.-h.	FDCf-II	$> 1/2$	1	ν	0	0	1	1
Fin.-h.	FDCc-II	$> 1/4$	1	2ν	0	0	1	1
Fin.-h.	IPA, LR		1	∞	0	0	1	1
Trunc.-h.	FDf	$1/4$	1	$1/4$	$1/2$	1	$1/2$	$1/4$
Trunc.-h.	FDc	$1/6$	1	$1/3$	$2/3$	1	$2/3$	$1/3$
Trunc.-h.	FDf	$1/3$	1	$1/3$	$-1/3$	$1/3$	$2/3$	$1/2$
Trunc.-h.	FDc	$1/4$	1	$1/2$	0	$1/2$	1	$2/3$
Trunc.-h.	FDCf-I	$(p+1)/3$	1	ν	$p-\nu$	$\geq 1/2$	2ν	$2/3$
Trunc.-h.	FDCc-I	$(p+1)/5$	1	2ν	$p-\nu$	$\geq 2/3$	4ν	$4/5$
Trunc.-h.	FDCf-II	$\geq (p+1)/2$	1	$\nu \wedge p$	p	≥ 1	$p+1$	1
Trunc.-h.	FDCc-II	$\geq (p+1)/2$	1	$(2\nu) \wedge p$	p	≥ 1	$p+1$	1
Trunc.-h.	IPA		1	p	p	≥ 1	$p+1$	1
Trunc.-h.	LR		1	$1/3$	$-1/3$	$1/3$	$2/3$	$1/3$
Trunc.-h.	CLR		1	$1/2$	0	$1/2$	1	$2/3$

$$c_n = O(n^{-\nu}); \qquad \gamma_n = O(n^{-\gamma});$$

$$t_n = O(n^p); \qquad C_n = O(n^{p+1});$$

$$B_n = O(n^{-\beta}) = O(n^{-bp}); \qquad V_n = O(n^{-\delta}) = O(n^{-dp});$$

$$E[(\theta_n - \theta_*)^2] = O(n^{-\kappa}) = O(C_n^{-\eta}).$$

Averaging.

SA is very sensitive to the choice of γ_0.

Optimal γ_0 is $\gamma_* = [\alpha''(\theta_*)]^{-1}$, but unknown.

Recently proposed averaging methods are much less sensitive and converge as fast, provided that $1/2 < \gamma < 1$.

1. Post-averaging (Polyak 1990; Kushner and Yang 1993):

$$\begin{aligned}
\theta_{n+1} &= \theta_n - n^{-\gamma} Y_n; \\
\bar{\theta}_{n+1} &= \frac{1}{n+1} \sum_{j=1}^{n+1} \theta_j = \bar{\theta}_n + \frac{1}{n+1}(\theta_{n+1} - \bar{\theta}_n).
\end{aligned}$$

2. Double-averaging (Bather 1989; Yin and Yin 1994):

$$\begin{aligned}
\theta_{n+1} &= \bar{\theta}_n - n^{-\gamma} \sum_{j=1}^{n} Y_j; \\
\bar{\theta}_{n+1} &= \frac{1}{n+1} \sum_{j=1}^{n+1} \theta_j = \bar{\theta}_n + \frac{1}{n+1}(\theta_{n+1} - \bar{\theta}_n).
\end{aligned}$$

$GI/GI/1$ toy example (continued): optimization.

Want to minimize $E\left[\frac{1}{t}\sum_{i=1}^{t} X_i\right] + K/\theta$.

SC: compute $H'_N(\cdot)$ and solve $H'_N(\theta) - K/\theta^2 = 0$.

SA: take $Y_n = H'_n(\theta_n, \omega_n) - K/\theta_n^2$.

Numerical results for M/M/1 queue.

$\lambda = 1$, $K = 1$, $\Theta = [0.1, 0.9]$, $t = 10$, $\theta_* = .5965$.

Table 1: SA results for $R = 10$ ($\theta_* = .5965$).

	γ_0	$\theta_0 - \theta_n$	MSE	
			$N = 10^4$	$N = 10^5$
IPA	0.1	0.1	4.5E-7	4.2E-8
IPA	0.5	0.1	2.1E-6	2.8E-7
IPA	0.1	0.0	2.6E-6	2.8E-7
SF	0.1	0.2	1.0E-5	1.7E-6
SF	0.1	0.0	4.6E-5	6.6E-6

Table 2: SC results for $R = 10$ ($\theta_* = .5965$).

	θ_0	MSE
		$N = 10^4$
IPA	0.7	3.8E-7
IPA	0.6	2.7E-6
SF	0.8	8.4E-6
SF	0.6	6.2E-5

$GI/GI/1$ example (continued): steady-state optimiz.

Want to minimize $w(\theta) + K/\theta$, where

$w(\theta)$ = average infinite-horizon waiting time $= u(\theta)/\ell(\theta)$;

$$\ell(\theta) = E[\tau] \qquad\qquad u(\theta) = E\left[\sum_{i=1}^{\tau} W_i\right].$$

Want to solve

$$\ell(\theta)u'(\theta) - u(\theta)\ell'(\theta) - K\ell(\theta)\ell(\theta)/\theta^2 \; = \; 0. \tag{1}$$

SC algorithm:

FOR $n := 1$ TO N DO

From one regen. cycle, estimate ℓ and u;

From another (indep.), estimate ℓ, ℓ', and u';

This gives an unbiased estim. H'_n of l.h.s. of (1).

Compute $\bar{H}'_N(\cdot)$ and solve $\bar{H}'_N(\theta) = 0$.

For SA algorithm: can take $Y_n = H'_n(\theta_n)$.

Numerical results for M/M/1 queue.

$\lambda = 1$, $K = 1$, $\Theta = [0.1, 0.9]$, $\theta_* = 0.5$,

Here, N is random; we stop after simulating a grand total of C customers.

Table 3: SA results for $R = 10$ ($\theta_* = 0.5$).

	γ_0	$\theta_0 - \theta_n$	MSE	
			$C = 10^5$	$C = 10^6$
IPA-R	0.03	.00	9.0E-7	4.5E-7
IPA-R	0.03	.10	4.2E-6	4.3E-7
IPA-R	0.03	.20	2.1E-6	5.8E-7
IPA-R	0.03	.30	1.2E-5	2.2E-6
SF-R	0.01	.00	1.1E-4	1.3E-5
SF-R	0.01	.10	4.0E-5	2.3E-6
SF-R	0.01	.20	6.6E-5	6.9E-6
SF-R	0.01	.30	1.1E-4	7.3E-6

Table 4: SC results for $R = 10$ ($\theta_* = 0.5$).

	θ_0	MSE	
		$C = 10^5$	$C = 10^6$
IPA-R	0.5	1.00E-6	4.90E-7
IPA-R	0.6	1.00E-6	4.90E-7
IPA-R	0.7	1.00E-6	4.90E-7
IPA-R	0.8	1.00E-6	4.90E-7
SF-R	0.5	1.00E-5	1.00E-6
SF-R	0.6	1.00E-5	1.00E-6
SF-R	0.7	1.00E-5	1.00E-6
SF-R	0.8	1.00E-5	1.00E-6

Other

— FD with simultaneous perturbations (Spall 1992).

— Frequency domain experimentation (Schruben and Cogliano, Morrice, Jacobson, Sargent and Som, ...).

— Functional estimation without LR (Vázquez-Abad and L'Ecuyer 1994).

— Other PA variants (phantom method, etc.).

— Examples which are "open" problems for efficient gradient estimation (e.g., multicomponent replacement problem ...).

— Global optimization (e.g., Yakowitz 1993).