Resit Optimization under Uncertainty 4.2

22 March 2018, 12:00-14:45h

Problem 1 (10 Credits)

Let $J \in \mathcal{C}^2$ be such that the gradient is a bounded and Lipschitz continuous function and consider the fixed ϵ algorithm:

$$\theta_{n+1} = \theta_n - \epsilon \nabla_\theta J(\theta_n). \tag{1}$$

Suppose that θ_n converges to some finite admissible θ^* as n tends to infinity. Show that θ^* is a stationary point of J.

Answer Problem 1: Convergence of θ_n to θ^* implies

$$\lim_{n\to\infty} \nabla_\theta J(\theta_n) = 0.$$

We have assumed that $\nabla_\theta J$ is Lipschitz continuous and thus continuous, which gives

$$0 = \lim_{n\to\infty} \nabla_\theta J(\theta_n) = \nabla_\theta J\left(\lim_{n\to\infty} \theta_n\right) = \nabla_\theta J(\theta^*),$$

and we conclude that θ^* is a stationary point.

Problem 2 (total 30 Credits)

Consider the following reservoir model. Per time period the amount of inflowing fluid is I_t and the amount of outflowing liquid is O_t. Let $L_t \geq 0$ denote the level of the fluid in the reservoir at the end of the t-th time period. Then,

$$L_{t+1} = \max(L_t + I_t - O_t, 0).$$

Assume that $I_t = I_t(\theta)$ follows a $\mathsf{Gamma}(2, \theta^{-1})$ distribution, i.e., I_t behaves like the sum of two independent exponentially distributed random variables with mean θ each, and that O_t is log-normal distributed independent of I_t. The cost for having an inflow at θ, is given by a θ^{-2}. Let

$$J(\theta) = \mathbb{E}[L_{t_0+1} | L_{t_0} = l] + \theta^{-2},$$

for some $t_0 > 2$ and $l > 0$, and consider the problem

$$\min_\theta J(\theta).$$

In words, given the reservoir level is $l > 0$, we want to regulate the inflow so that the expected reservoir level at the end of the next time period is minimal.

You may assume that there is a unique stationary point to the function $J(\theta)$ that provides the location of the minimum.

(a). [5 Credits] Compute the IPA estimator for $\nabla J(\theta)$ (you don't have to check unbiasedness).

(b). [5 Credits] Using the IPA estimator from (a) provide a descent algorithm for finding the solution of the minimization problem.

(c). [5 Credits] Letting $\epsilon_n = 1/(n+1)$, what properties have to checked for establishing a.s. convergence of your algorithm to the location of the minimum?

(d). [15 Credits] Now assume that you are interested adjusting the fluid level to α, that is, you want to find θ^* such that

$$\mathbb{E}[L_{t_0+1}|L_{t_0} = l] = \alpha.$$

For this exercise we keep l and t_0 fixed and we use simulation to find the "right" θ^* for period $t_0 + 1$. Provide a descent algorithm and discuss sufficient condition for its convergence to θ^*. You may assume that the solution θ^* is asymptotically stable for your vector field $G(\theta)$ (which you will provide) and that $G(\theta)$ is continuous and bounded. Moreover, assume that $\mathrm{Var}(\mathbb{E}[L_{t_0+1}|L_{t_0} = l]) \leq c$ for all n.

Answer Problem 2: (a) By assumption $I_t = I_t(\theta)$ follows a $\mathsf{Gamma}(2, \theta^{-1})$ distribution, therefore we may let

$$I_t(\theta) = X_1(\theta) + X_2(\theta),$$

where $X_i(\theta)$ are independent exponential with mean θ. Then,

$$\frac{d}{d\theta} I_t(\theta) = \frac{d}{d\theta}(X_1(\theta) + X_2(\theta)) = \frac{1}{\theta} I_t(\theta).$$

Under the condition that $L_{t_0} = l$, the IPA estimator becomes

$$\frac{d}{d\theta} L_{t_0+1} = \frac{1}{\theta} I_t(\theta) 1_{l+I_t(\theta)-O_t \geq 0} - 2\theta^{-3}.$$

(b) Let

$$Y_n = -\left(\frac{1}{\theta} I_t(\theta) 1_{l+I_t(\theta)-O_t \geq 0} - 2\theta^{-3}\right),$$

then

$$\theta_{n+1} = \theta_n - \frac{1}{n+1}\left(\frac{1}{\theta} I_t(\theta) 1_{l+I_t(\theta)-O_t \geq 0} - 2\theta^{-3}\right).$$

(c) The conditions to be checked are (i) unbiasedness of the algorithm, i.e.,

$$\mathbb{E}[Y_n|\mathcal{F}_{n-1}] = \nabla J(\theta_n)$$

and the variance condition

$$\sum_{n=1}^{\infty} \frac{1}{n+1} \mathbb{E}\big[\big(-\big(\frac{1}{\theta} I_t(\theta) 1_{l+I_t(\theta)-O_t \geq 0} - 2\theta^{-3}\big) + \nabla J(\theta_n)\big)^2 \big| \mathcal{F}_{n-1}\big] < \infty.$$

(d) Let

$$G(\theta) = \alpha - \mathbb{E}[L_{t_0+1}|L_{t_0} = l].$$

Since $\mathbb{E}[L_{t_0+1}|L_{t_0} = l]$ is monotone increasing, point θ^* is for the ODE

$$\frac{d}{dt}x(t) = \alpha - \mathbb{E}[L_{t_0+1}|L_{t_0} = l] = G(x(t))$$

asymptotically stable. We thus consider

$$\theta_{n+1} = \theta_n + \epsilon_n(\alpha - \max(l + I_t(\theta) - O_t, 0)).$$

We let $\epsilon_n = 1/n$. $Y_n = \alpha - \max(l + I_t(\theta_n) - O_t, 0)$ is unbiased for $G(\theta_n)$ and it remains to check the variance condition. As usual,

$$V_n = \mathbb{E}\left[(\alpha - \max(l + I_t(\theta_n) - O_t, 0) - G(\theta_n))^2 \middle| \mathcal{F}_{n-1}\right] = \mathrm{Var}(\max(l + I_t(\theta_n) - O_t, 0)).$$

As we have assumed that the variance is uniformly bounded by some constant c and we compute as usual

$$\sum_n \epsilon_n^2 V_n = \sum_n \epsilon_n^2 \mathrm{Var}(\max(l + I_t(\theta_n) - O_t, 0)) \leq c \sum_n \epsilon_n^2 < \infty.$$

Problem 3 (total 10 Credits)

The gradient-field of a function $J(\theta)$ is shown in Figure 1. Apply a steepest descent algorithm for

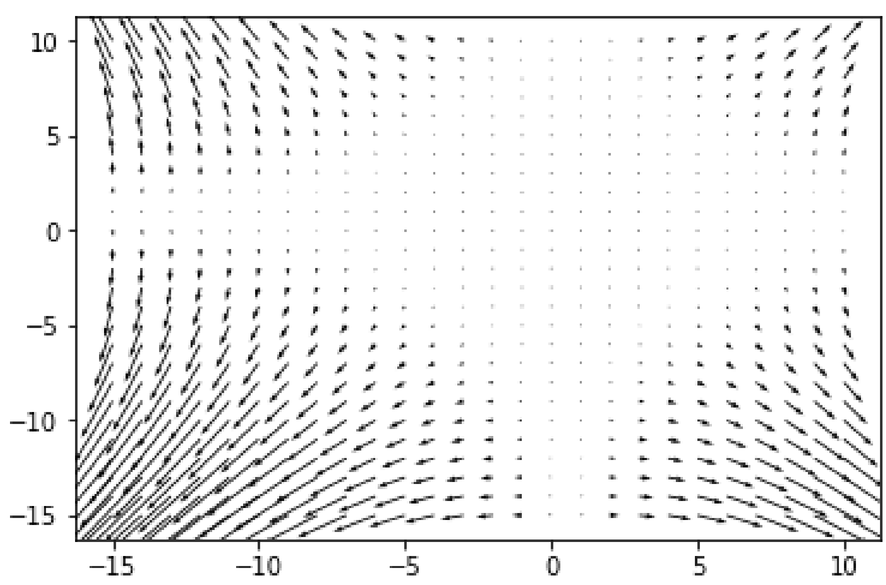

Figure 1: Gradient-field of $J(\theta)$

finding the minimum of $J(\theta)$.

(a). [5 Credits] If you choose point $(-15,-10)$ as initial point. Argue with Figure 1 that the algorithm will not converge to $(0,0)$.

(b). [5 Credits] For the same ODE as in part (a) discuss the nature of point $(0,0)$ (stable, asymptotically stable, or unstable).

Answer Problem 3: (a) The graph shows the gradients, so the negative gradients point in opposite direction. Starting in $(-15,-10)$ the ODE will be drawn towards a point near $(-13,0)$. Hence, the ODE will not reach $(0,0)$.

(b) The point is stable as the gradient is zero in this point. The point is not asymptotically stable as the gradient around $(0,0)$ is (almost) zero. [The ODE will not move in the neighborhood of $(0,0)$]

Problem 4 (total 50 Credits)

Let $X = X_\theta \sim \mathsf{Exp}(\theta^{-1})$ for $\theta > 0$ (whatever convenient we write X or X_θ). The pdf of X is

$$f_\theta(x) = \frac{1}{\theta}\, e^{-x/\theta}, \quad x \geq 0.$$

In the sequel you may use that for any power $p \in \mathbb{N}$, $\mathbb{E}\big[X_\theta^p\big] = p!\,\theta^p$. Let $J(\theta) = \mathbb{E}\big[X_\theta^2\big]$. In this problem you are going to analyse unbiased estimators of $J'(\theta) = \frac{d}{d\theta}J(\theta)$.

(a). [15 Credits] Define

$$D = D_\theta = \frac{2X_\theta^2}{\theta}.$$

(i). Derive that D is the IPA estimator of $J'(\theta)$, assuming that the IPA interchange condition hold, i.e.,

$$\frac{d}{d\theta}\mathbb{E}\big[X_\theta^2\big] = \mathbb{E}\big[\frac{d}{d\theta}X_\theta^2\big]. \tag{2}$$

(ii). Show that D is unbiased.

(iii). Compute the variance of D (answer is $80\theta^2$).

(iv). Argue that the interchange condition (2) hold.

(b). [15 Credits] Define

$$D = D_\theta = \frac{X_\theta^2}{\theta}\Big(\frac{X_\theta}{\theta} - 1\Big).$$

(i). Derive that D is the SFM estimator of $J'(\theta)$, assuming that the SFM interchange condition hold, i.e.,

$$\frac{d}{d\theta}\int x^2 f_\theta(x)\, dx = \int x^2 \frac{d}{d\theta} f_\theta(x)\, dx. \tag{3}$$

(ii). Show that D is unbiased.

(iii). Compute the variance of D (answer is $448\theta^2$). You may use that $\mathrm{Cov}\big(X^3/\theta^2, X^2/\theta\big) = 108\theta^2$.

(iv). Argue that the interchange condition (3) hold.

(c). [15 Credits] Define

$$D = D_\theta = \frac{1}{\theta}\Big((X_\theta + Y_\theta)^2 - X_\theta^2\Big),$$

where $Y = Y_\theta$ is independent of X_θ, and also $\mathsf{Exp}(\theta^{-1})$ distributed.

(i). Derive that D is the MVD estimator of $J'(\theta)$. Namely, assume that the interchange condition (3) hold, then derive that

$$\frac{d}{d\theta} f_\theta(x) = \frac{1}{\theta}\big(g_\theta(x) - f_\theta(x)\big),$$

with $g_\theta(x)$ the pdf of $\mathsf{Gamma}(2, \theta^{-1})$, i.e.,

$$g_\theta(x) = \frac{x}{\theta^2} e^{-x/\theta}, \quad x \geq 0.$$

(ii). Show that D is unbiased. Hint: $(X+Y)^2 = X^2 + 2XY + Y^2$ and X, Y are independent.

(iii). Compute the variance of D (answer is $48\theta^2$). You may use that

$$\mathbb{E}\big[(X+Y)^4\big] = 120\theta^4; \text{ and } \mathrm{Cov}\big((X+Y)^2, X^2\big) = 28\theta^4.$$

(d). [5 Credits] What is your conclusion?

Answers Problem 4:

(a). Note $J(\theta) = \mathbb{E}[h(X_\theta)]$ with $h(x) = x^2$, and by the inverse transform method, $X_\theta = -\theta \ln(1-$ U where U is uniform (0,1):

$$F_\theta(x) = 1 - e^{-x/\theta} = u \quad \Leftrightarrow \quad x = -\theta \ln(1-u).$$

Thus by the chain rule:

$$\frac{\partial}{\partial\theta} h(X_\theta) = h'(X_\theta) X_\theta' = -2X_\theta \ln(1-U) = \frac{-2X_\theta \theta \ln(1-U)}{\theta} = \frac{2X_\theta^2}{\theta}.$$

The interchange (2) is

$$J'(\theta) = \frac{\partial}{\partial\theta}\mathbb{E}[h(X_\theta)] = \mathbb{E}\big[\frac{\partial}{\partial\theta} h(X_\theta)\big],$$

which shows the IPA estimator

$$D_\theta = \frac{\partial}{\partial\theta} h(X_\theta) = \frac{2X_\theta^2}{\theta}.$$

To show unbiasedness for $J'(\theta)$, we use $\mathbb{E}\big[X_\theta^p\big] = p!\,\theta^p$. Firstly,

$$J(\theta) = \mathbb{E}[X_\theta^2] = 2\theta^2 \quad \Rightarrow \quad J'(\theta) = 4\theta.$$

Next,

$$\mathbb{E}[D_\theta] = \frac{2}{\theta}\mathbb{E}[X_\theta^2] = \frac{2}{\theta}\,2!\theta^2 = 4\theta = J'(\theta).$$

For the variance, compute the second moment:

$$\mathbb{E}[D_\theta^2] = \frac{4}{\theta^2}\mathbb{E}[X_\theta^4] = \frac{4}{\theta^2}\,4!\theta^4 = 96\theta^2.$$

Thus

$$\mathbb{V}ar[D_\theta] = \mathbb{E}[D_\theta^2] - \big(\mathbb{E}[D_\theta]\big)^2 = 96\theta^2 - 16\theta^2 = 80\theta^2.$$

Interchange is allowed because (i) X_θ is differentiable (in θ), (ii) $h(x)$ is differentiable (in x), and (iii) $Y(\theta) \doteq h(X(\theta))$ is almost surely Lipschitz continuous on any interval $(a,b) \subset (0,\infty)$. To show condition (iii):

$$\begin{aligned}\sup_{\theta\in(a,b)} |Y'(\theta)| &= \sup_{\theta\in(a,b)} \frac{2X_\theta^2}{\theta} \\ &= \sup_{\theta\in(a,b)} \frac{2\theta^2}{\theta}\big(\ln(1-U)\big)^2 = 2b\big(\ln(1-U)\big)^2 < \infty.\end{aligned}$$

The Lipschitz modulus is

$$K = 2b\big(\ln(1-U)\big)^2 = \big(\underbrace{-\sqrt{2b}\,\ln(1-U)}_{\overset{\mathcal{L}}{\sim}\mathsf{Exp}(1/\sqrt{2b})}\big)^2 \;\Rightarrow\; \mathbb{E}[K] < \infty.$$

(b). The pdf of X_θ is $f_\theta(x) = \frac{1}{\theta}\,e^{-x/\theta}$, which gives the score function:

$$S(\theta,x) \doteq \frac{\partial}{\partial\theta}\ln f_\theta(x) = \frac{\partial}{\partial\theta}\big(-\ln\theta - \frac{x}{\theta}\big) = -\frac{1}{\theta} + \frac{x}{\theta^2}.$$

Work out the interchange (3):

$$\begin{aligned}J'(\theta) &= \frac{\partial}{\partial\theta}\mathbb{E}[X_\theta^2] = \frac{\partial}{\partial\theta}\int x^2 f_\theta(x)\,dx = \int x^2\frac{\partial}{\partial\theta}f_\theta(x)\,dx = \int x^2\frac{\frac{\partial}{\partial\theta}f_\theta(x)}{f_\theta(x)}\,f_\theta(x)\,dx \\ &= \int x^2\frac{\partial}{\partial\theta}\big(\ln f_\theta(x)\big)\,f_\theta(x)\,dx = \int x^2 s(\theta,x)\,f_\theta(x)\,dx = \mathbb{E}\big[X_\theta^2\,S(\theta,X_\theta)\big].\end{aligned}$$

This shows the SFM estimator

$$D_\theta = X_\theta^2\,S(\theta,X_\theta) = X_\theta^2\,\big(-\frac{1}{\theta} + \frac{X_\theta}{\theta^2}\big) = \frac{X_\theta^2}{\theta}\big(\frac{X_\theta}{\theta} - 1\big) = \frac{X_\theta^3}{\theta^2} - \frac{X_\theta^2}{\theta}.$$

Recall that $J'(\theta) = 4\theta$. Then

$$\mathbb{E}[D_\theta] = \mathbb{E}\big[\frac{X_\theta^3}{\theta^2} - \frac{X_\theta^2}{\theta}\big] = \frac{3!\theta^3}{\theta^2} - \frac{2!\theta^2}{\theta} = 6\theta - 2\theta = 4\theta.$$

The variance:

$$\mathbb{V}ar[D_\theta] = \mathbb{V}ar\big[\frac{X_\theta^3}{\theta^2} - \frac{X_\theta^2}{\theta}\big] = \mathbb{V}ar\big[\frac{X_\theta^3}{\theta^2}\big] + \mathbb{V}ar\big[\frac{X_\theta^2}{\theta}\big] - 2\mathbb{C}ov\big(\frac{X_\theta^3}{\theta^2},\frac{X_\theta^2}{\theta}\big).$$

Work out the three terms:

$$\begin{aligned}\mathbb{V}ar\big[\frac{X_\theta^3}{\theta^2}\big] &= \frac{1}{\theta^4}\,(6!\theta^6 - (3!\theta^3)^2) = 720\theta^2 - 36\theta^2 = 684\theta^2 \\ \mathbb{V}ar\big[\frac{X_\theta^2}{\theta}\big] &= \frac{1}{\theta^2}\,(4!\theta^4 - (2!\theta^2)^2) = 24\theta^2 - 4\theta^2 = 20\theta^2 \\ 2\mathbb{C}ov\big(\frac{X_\theta^3}{\theta^2},\frac{X_\theta^2}{\theta}\big) &= 216\theta^2\end{aligned}$$

Which gives

$$\mathbb{V}ar[D_\theta] = 684\theta^2 + 20\theta^2 - 216\theta^2 = 488\theta^2.$$

The nontrivial interchange condition is

$$\int x^2 \sup_{\theta\in(a,b)} \big|\frac{\partial}{\partial\theta} f_\theta(x)\big|\, dx < \infty.$$

Work out,

$$\frac{\partial}{\partial\theta} f_\theta(x) = \big(\frac{x}{\theta^3} - \frac{1}{\theta^2}\big)\, e^{-x/\theta} = \frac{e^{-x/\theta}}{\theta^2}\big(\frac{x}{\theta} - 1\big). \tag{4}$$

Thus, for $a < \theta < b$ and $x > a$ is

$$\frac{e^{-x/\theta}}{\theta^2} \le \frac{e^{-x/b}}{a^2}; \text{ and } \big|\frac{x}{\theta} - 1\big| \le \frac{x}{a}.$$

Hence,

$$\int_a^\infty x^2 \sup_{\theta\in(a,b)} \big|\frac{\partial}{\partial\theta} f_\theta(x)\big|\, dx \le \int_a^\infty \frac{x^3}{a^3} e^{-x/b}\, dx < \infty.$$

(c). Differentiate $f_\theta(x) = e^{-x/\theta}/\theta$, see (4):

$$\frac{\partial}{\partial\theta} f_\theta(x) = \frac{1}{\theta}\Big(\underbrace{\frac{x}{\theta^2} e^{-x/\theta}}_{=g_\theta(x) \stackrel{\mathcal{L}}{=} \mathsf{Gamma}(2,1/\theta)} - \underbrace{\frac{1}{\theta} e^{-x/\theta}}_{=f_\theta(x) \stackrel{\mathcal{L}}{=} \mathsf{Exp}(1/\theta)} \Big).$$

Because the $\mathsf{Gamma}(2,\alpha)$ is the sum of two iid $\mathsf{Exp}(\alpha)$ random variables, we let $X_\theta, Y_\theta \stackrel{\mathcal{L}}{\sim} \mathsf{Exp}(1/\theta)$ independent, and $X_\theta + Y_\theta \stackrel{\mathcal{L}}{\sim} \mathsf{Gamma}(2, 1/\theta)$. This gives that

$$\begin{aligned}
J'(\theta) &= \frac{\partial}{\partial\theta}\mathbb{E}[X_\theta^2] = \frac{\partial}{\partial\theta}\int x^2 f_\theta(x)\, dx = \int x^2 \frac{\partial}{\partial\theta} f_\theta(x)\, dx \\
&= \int x^2 \frac{1}{\theta}\big(g_\theta(x) - f_\theta(x)\big)\, dx = \frac{1}{\theta}\big(\int x^2 g_\theta(x)\, dx - \int x^2 f_\theta(x)\, dx\big) \\
&= \frac{1}{\theta}\big(\mathbb{E}\big[(X_\theta + Y_\theta)^2\big] - \mathbb{E}\big[X_\theta^2\big]\big) = \frac{1}{\theta}\mathbb{E}\big(\big[(X_\theta + Y_\theta)^2\big] - X_\theta^2\big)\big].
\end{aligned}$$

This shows the MVD estimator

$$D_\theta = \frac{1}{\theta}\Big((X_\theta + Y_\theta)^2 - X_\theta^2\Big).$$

For unbiasedness, recall that $J'(\theta) = 4\theta$. Then

$$\theta\mathbb{E}[D_\theta] = \mathbb{E}\big[(X_\theta + Y_\theta)^2 - X_\theta^2\big]\mathbb{E}\big[(X_\theta + Y_\theta)^2\big] - \mathbb{E}\big[X_\theta^2\big],$$

with

$$\begin{aligned}
&\mathbb{E}\big[(X_\theta + Y_\theta)^2\big] = \mathbb{E}\big[X_\theta^2 + Y_\theta^2 + 2X_\theta Y_\theta\big] = \mathbb{E}[X_\theta^2] + \mathbb{E}[Y_\theta^2] + 2\mathbb{E}[X_\theta]\mathbb{E}[Y_\theta] \\
&= 2\theta^2 + 2\theta^2 + 2\theta^2 = 6\theta^2 \\
&\mathbb{E}\big[X_\theta^2\big] = 2\theta^2.
\end{aligned}$$

Thus $\mathbb{E}[D_\theta] = 6\theta - 2\theta = 4\theta$. The variance:

$$\theta^2 \mathbb{V}ar[D_\theta] = \mathbb{V}ar\big[(X_\theta + Y_\theta)^2 - X_\theta^2\big] = \mathbb{V}ar\big[(X_\theta + Y_\theta)^2\big] + \mathbb{V}ar\big[X_\theta^2\big] - 2\mathbb{C}ov\big((X_\theta + Y_\theta)^2, X_\theta^2\big).$$

Work out the three terms:

$$\begin{aligned}
&\mathbb{V}ar\big[(X_\theta + Y_\theta)^2\big] = \mathbb{E}\big[(X_\theta + Y_\theta)^4\big] - \big(\mathbb{E}\big[(X_\theta + Y_\theta)^2\big]\big)^2 = 120\theta^4 - 36\theta^4 = 84\theta^4 \\
&\mathbb{V}ar\big[X_\theta^2\big] == \mathbb{E}\big[X_\theta^4\big] - \big(\mathbb{E}\big[X_\theta^2\big]\big)^2 = 4!\theta^4 - (2!\theta^2)^2 = 20\theta^4 \\
&2\mathbb{C}ov\big((X_\theta + Y_\theta)^2, X_\theta^2\big) = 56\theta^4.
\end{aligned}$$

Which gives

$$\mathbb{V}ar[D_\theta] = 84\theta^2 + 20\theta^2 - 56\theta^2 = 48\theta^2.$$

(d). The conclusion is that we see here three unbiased estimators of $J'(\theta)$, but their variances differ quite a bit. In fact,

$$\mathbb{V}ar\big[D_\theta^{\text{MVD}}\big] \leq \mathbb{V}ar\big[D_\theta^{\text{IPA}}\big] \leq \mathbb{V}ar\big[D_\theta^{\text{SFM}}\big].$$

On the other hand, when computing these estimator by simulation, the MVD estimator takes double computation times.