Liam Hardiman *May 27, 2020*

270C - Homework 3

5.1.2

Prove the following statements.

(a) Every Lipschitz function is uniformly continuous.

Proof. Let $f : X \to Y$ be a Lipschitz function between metric spaces (X, d_X) and (Y, d_Y) with Lipschitz constant K. Let $\epsilon > 0$ be given. Then if $d_X(x, y) < \epsilon/K$, we have that

$$d_Y(f(x), f(y)) \leq K d_X(x, y) \leq \epsilon,$$

so f is uniformly continuous. $\square$

(b) Every differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz and

$$\|f\|_{\text{Lip}} \leq \sup_{x \in \mathbb{R}^n} \|\nabla f(x)\|_2.$$

Proof. Let $\epsilon > 0$ be given. Since f is differentiable, for each x there exists a δ_x such that $\|x - y\| < \delta$ implies that

$$\|f(y) - f(x) - \nabla f(x)^T (y - x)\|_2 < \epsilon \|y - x\|_2.$$

By the triangle inequality, $\|f(y) - f(x)\|_2 \leq (\|\nabla f(x)\|_2 + \epsilon)\|y - x\|_2$. Taking $\epsilon \to 0$ establishes the bound (locally). If we restrict ourselves to a compact set (gets rid of locality) and assume f is continuously differentiable we can establish the desired claim I think. $\square$

(c) Give an example of a non-Lipshitz but uniformly continuous function $f : [-1, 1] \to \mathbb{R}$.

Proof. Consider $f(x) = |x|^{1/2}$. Qualitatively, the sharp cusp at the origin keeps f from being Lipshitz on $[-1, 1]$, while f is uniformly continuous since it's a continuous function on a compact set. More rigorously, $|x|^{1/2}/|x| \to \infty$ as $x \to 0$, so there is no K such that $|x|^{1/2} \leq K|x|$ for all x sufficiently small. $\square$

(d) Give an example of a non-differentiable but Lipshitz function $f : [-1, 1] \to \mathbb{R}$.

Proof. Since Lipshitz functions are differentiable almost everywhere, the most exotic thing we can hope for is something with only a few bad points. For example, $f(x) = |x|$ is as Lipshitz as it gets, but it isn't differentiable at the origin. $\square$

5.1.3

Prove the following statements.

(a) For a fixed $\theta \in \mathbb{R}^n$, the linear functional

$$f(x) = \langle x, \theta \rangle$$

is a Lipshitz function on $\mathbb{R}^n$ and $\|f\|_{\text{Lip}} = \|\theta\|_2$.

Proof. That f is Lipshitz follows from Cauchy-Schwartz:

$$|f(x) - f(y)| = |\langle x - y, \theta \rangle| \le \|x - y\|_2 \cdot \|\theta\|_2.$$

In particular, $\|f\|_{\text{Lip}} \le \|\theta\|_2$. Now we also have

$$\frac{|f(x) - f(y)|}{\|x - y\|_2} = \frac{|\langle x - y, \theta \rangle|}{\|x - y\|_2} = |\cos \phi| \cdot \|\theta\|_2,$$

where ϕ is the angle between $x - y$ and θ. Since $|\cos \phi|$ can reach its maximum value of 1, we conclude that the Lipschitz norm of f is indeed $\|\theta\|_2$. □

(b) More generally, an $m \times n$ matrix A acting as a linear operator

$$A : (\mathbb{R}^n, \|\cdot\|_2) \to (\mathbb{R}^m, \|\cdot\|_2)$$

is Lipschitz and $\|A\|_{\text{Lip}} = \|A\|$.

Proof. A is certainly Lipschitz since for any $x, y \in \mathbb{R}^n$,

$$\|Ax - Ay\|_2 = \|A(x - y)\|_2 \le \|A\| \cdot \|x - y\|_2.$$

In particular, $\|A\|_{\text{Lip}} \le \|A\|$. Furthermore, by linearity we have

$$\|A\|_{\text{Lip}} = \inf \left\{ K : \frac{\|A(x - y)\|_2}{\|x - y\|_2} \le K, \text{ for all } x, y \in \mathbb{R}^n \right\} = \sup_{v \in \mathbb{R}^n,\ v \neq 0} \frac{\|Av\|_2}{\|v\|_2} = \|A\|.$$

□

(c) Any norm $f(x) = \|x\|$ on $(\mathbb{R}^n, \|\cdot\|_2)$ is a Lipschitz function. The Lipschitz norm of f is the smallest L that satisfies

$$\|x\| \le L\|x\|_2 \quad \text{for all } x \in \mathbb{R}^n.$$

Proof. It's a standard fact from real analysis that all norms on $\mathbb{R}^n$ are equivalent, so there are positive constants C_1, C_2 such that

$$C_1\|x\|_2 \le \|x\| \le C_2\|x\|_2$$

for all $x \in \mathbb{R}^n$. By the reverse triangle inequality we then have

$$|f(x) - f(y)| = \big| \|x\| - \|y\| \big| \le \|x - y\| \le C_2\|x - y\|_2,$$

so f is Lipschitz. The same argument used in part (b) gives the Lipschitz constant. □

5.1.9

Let A be a subset of the sphere $\sqrt{n}S^{n-1}$ such that

$$\sigma(A) > 2\exp(-cs^2)$$

for some $s > 0$.

(a) Prove that $\sigma(A_s) > 1/2$.

Proof. Suppose, for the sake of contradiction, that $\sigma(A_s) \leq 1/2$. Then the complement $B := (A_s)^C$ satisfies $\sigma(B) \geq 1/2$ and we have by the blowup lemma that

$$\sigma(B_t) \geq 1 - 2\exp(-ct^2)$$

for all $t \geq 0$. By construction, the sets B_s and A are disjoint, but $\sigma(B_s) \geq 1 - 2\exp(-cs^2)$ and $\sigma(A) > 2\exp(-cs^2)$, a contradiction. □

(b) Deduce that for any $t \geq s$,

$$\sigma(A_{2t}) \geq 1 - 2\exp(-ct^2).$$

Proof. By part (a) and the blowup lemma we have that

$$\sigma\big((A_s)_t\big) \geq 1 - 2\exp(-ct^2)$$

for all $t \geq 0$. Setting $t \geq s$ gives the desired result since for such t we have $A_{2t} \supseteq (A_s)_t$. □

5.1.15

Fix $\epsilon \in (0, 1)$. Show that there exists a set $\{x_1, \ldots, x_N\}$ of unit vectors in $\mathbb{R}^n$ which are mutually almost orthogonal:

$$|\langle x_i, x_j \rangle| \leq \epsilon \quad \text{for all } i \neq j,$$

and the set is exponentially large in n:

$$N \geq \exp(c_\epsilon n).$$

Proof. Fix any x_0 in the sphere $\sqrt{n}S^{n-1}$ and consider the function $f_0(x) = \langle x, x_0 \rangle$. By exercise 5.1.3, f_0 is Lipschitz and $\|f_0\|_{\text{Lip}} = \|x_0\|_2 = \sqrt{n}$. If $X \sim Unif(\sqrt{n}S^{n-1})$, then by theorem 5.1.4 (concentration of Lipschitz functions on the sphere), we have that

$$\Pr\big[|f(X) - E[f(X)] > t\big] = \Pr[|\langle X, x_0 \rangle| > t] \leq 2\exp(-ct^2/n),$$

for all $t \geq 0$. Now we have that

$$\Pr\big[|\langle X, x_0 \rangle| > t\big] = \Pr\big[n|\cos\phi| > t\big],$$

where ϕ is the angle between X and x_0. Setting $t = n\epsilon$ gives

$$\Pr\left[|\cos\phi| > \epsilon\right] \leq 2\exp(-c\epsilon^2 n).$$

In particular, if we let E_0 be the set of vectors in $\sqrt{n}S^{n-1}$ that are almost orthogonal to x_0,

$$E_0 = \{x \in \sqrt{n}S^{n-1} : |\langle x, x_0\rangle| \leq \epsilon\},$$

then we have $\sigma(E_0) \geq 1 - 2\exp(-c\epsilon^2 n)$. This gives us a sort of algorithm for building our mutually almost orthogonal set. We first choose x_0 anywhere in $\sqrt{n}S^{n-1}$ and then choose $x_1 \in E_0$ as defined above. Then we choose x_2 in $E_0 \cap E_1$, where E_1 is defined analogously. A union bound says that on the k-th step we have

$$\sigma\left(\bigcap_{i=1}^{k} E_i\right) \geq 1 - 2k\exp(-c\epsilon^2 n).$$

The above quantity is strictly positive as long as $k < \frac{1}{2}\exp(c\epsilon^2 n)$, so we can build a family of mutually almost orthogonal vectors that is exponentially large in the dimension. $\square$

5.2.3

Consider a random vector $X \sim \mathcal{N}(0, I_n)$ and a Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ (with respect to the Euclidean metric). Show that

$$\|f(X) - Ef(X)\|_{\psi_2} \leq C\,\|f\|_{\text{Lip}}\,.$$

Proof. The proof strategy is the basically the same that we used to prove this result on the sphere. We use the Gaussian isoperimetric inequality, a blow-up argument for half spaces, and then use the same median argument to put it all together.

First suppose that $A \subseteq \mathbb{R}^n$ is measurable with Gaussian measure $\gamma(A) \geq 1/2$. Then we claim that for all $t \geq 0$,

$$\gamma(A_t) \geq 1 - \exp(-ct^2)$$

for some positive constant c. To see this, consider the half space, $H = \{x \in \mathbb{R}^n : x_1 \leq 0\}$. By assumption we have

$$\gamma(A) \geq \frac{1}{2} = \gamma(H).$$

Now by the Gaussian isoperimetric inequality, we have that $\gamma(A_t) \geq \gamma(H_t)$ for all $t \geq 0$. Now the t-neighborhood of a half space is again a half space:

$$H_t = \{x \in \mathbb{R}^n : x_1 \leq t\}.$$

Since $x_1 \sim \mathcal{N}(0, 1)$ and the standard normal is clearly sub-Gaussian, we have that

$$\gamma(H_t) \geq 1 - \exp(ct^2)$$

for some $c > 0$. From here, the proof is identical to the spherical case. $\square$

5.2.11

Let $\Phi(x)$ denote the cumulative distribution function of the standard normal distribution $\mathcal{N}(0,1)$. Consider a random vector $Z = (Z_1, \dots, Z_n) \sim \mathcal{N}(0, I_n)$. Show that

$$\phi(Z) := \big(\Phi(Z_1), \dots, \Phi(Z_n)\big) \sim \mathrm{Unif}([0,1]^n).$$

Proof. Since the coordinates of Z are independent, so are the coordinates of $\phi(Z)$. Since the coordinates of a $\mathrm{Unif}([0,1]^n)$ random variable are independent $\mathrm{Unif}([0,1])$ random variables, we just have to show that $\Phi(Z_i) \sim \mathrm{Unif}([0,1])$. This is clear since for all $t \in [0,1]$ we have.

$$\Pr[\Phi(Z_i) \le t] = \Pr[Z_i \le \Phi^{-1}(t)] = \Phi(\Phi^{-1}(t)) = t.$$

□

5.2.12

Let $X = \phi(Z)$ be as in the previous exercise. Use Gaussian concentration to control the deviation of $f(\phi(Z))$ in terms of $\|F \circ \phi\|_{\mathrm{Lip}} \le \|F\|_{\mathrm{Lip}} \|\phi\|_{\mathrm{Lip}}$. Show that $\|\phi\|_{\mathrm{Lip}}$ is bounded by an absolute constant and complete the proof of theorem 5.2.10.

Proof. By Gaussian concentration we have

$$\|(f \circ \phi)(Z) - E(f \circ \phi)(Z)\|_{\psi_2} \le C \|f \circ \phi\|_{\mathrm{Lip}} \le C \|f\|_{\mathrm{Lip}} \|\phi\|_{\mathrm{Lip}}.$$

Let's pick apart $\|\phi\|_{\mathrm{Lip}}$. The function Φ is differentiable and its derivative is bounded by 1. By an earlier exercise, we then have $|\Phi(x) - \Phi(y)| \le |x - y|$ for all $x, y \in \mathbb{R}$. This gives

$$\frac{\|\phi(x) - \phi(y)\|_2}{\|x - y\|_2} = \sqrt{\frac{\sum_{i=1}^n (\Phi(x_i) - \Phi(y_i))^2}{\sum_{i=1}^n (x_i - y_i)^2}} \le 1.$$

We then have that $\|\phi\|_{\mathrm{Lip}}$ is bounded by 1 and the claim follows. □

5.3.3

Let A be an $m \times n$ random matrix whose rows are independent, mean zero, sub-gaussian isotropic random vectors in $\mathbb{R}^n$. Show that the conclusion of the Johnson-Lindenstrauss lemma holds for $Q = (1/\sqrt{m})A$.

Proof. Our plan is to show that with probability at least $1 - 2\exp(-c\epsilon^2 m)$, Q satisfies

$$(1-\epsilon)\|x\|_2 \le \|Qx\|_2 \le (1+\epsilon)\|x\|_2 \quad \text{for all } x \in \mathbb{R}^n.$$

By linearity we can assume that $\|x\|_2 = 1$. We have that

$$(Qx)_i = \frac{1}{\sqrt{m}} \sum_{j=1}^n Q_{ij} x_j.$$

By our assumptions on Q, we have that $(Qx)_i$ is centered, has unit variance, and is sub-Gaussian. □

5.4.5

Prove the following properties.

(a) $\|X\| \leq t$ if and only if $-tI \preceq X \preceq tI$.

Proof. The eigenvalues of $X \pm tI$ are $\lambda \pm t$, where λ is an eigenvalue of X. Now $\|X\| = |\lambda_1|$, where λ_1 is the eigenvalue with the largest magnitude. If $\|X\| \leq t$, then $0 \leq \lambda + t$ and $t - \lambda \geq 0$ for all eigenvalues λ of X, so $-tI \preceq X \preceq tI$.

Suppose now that $-tI \preceq X \preceq tI$. By the same reasoning used above, every eigenvalue of X has magnitude less than t, so $\|X\| \leq t$. □

(b) Let $f, g : \mathbb{R} \to \mathbb{R}$ be two functions. If $f(x) \leq g(x)$ for all $x \in \mathbb{R}$ satisfying $|x| \leq K$, then $f(X) \preceq g(X)$ for all X satisfying $\|X\| \leq K$.

Proof. Suppose $\|X\| \leq K$. The eigenvalues of $g(X) - f(X)$ are $g(\lambda) - f(\lambda)$ where λ is an eigenvalue of X. Since $f(x) \leq g(x)$ for all $|x| \leq K$, we then have that the eigenvalues of $g(X) - f(X)$ are nonnegative, so $f(X) \preceq g(X)$. □

(c) Let $f : \mathbb{R} \to \mathbb{R}$ be an increasing function and X, Y are commuting matrices. Then $X \preceq Y$ implies $f(X) \preceq f(Y)$.

Proof. Since X and Y commute, they are simultaneously diagonalizable and the eigenvalues of the difference $Y - X$ are $\lambda_i - \mu_i$ for a particular ordering of the eigenvalues of Y and X. Since $X \preceq Y$, we have that $\lambda_i - \mu_i \geq 0$. Since f is increasing, $f(\lambda_i) - f(\mu_i) \geq 0$. These are the eigenvalues of $f(X) - f(Y)$, so $f(X) \preceq f(Y)$. □

(d) Give an example showing that property (c) may fail for non-commuting matrices.

Proof. The idea is to find two 2×2 matrices A, B such that $0 \preceq A \preceq B$ but $A^2 \npreceq B^2$. □

(e) Show that $X \preceq Y$ always implies $\mathrm{tr} f(X) \leq \mathrm{tr} f(Y)$ for any increasing function $f : \mathbb{R} \to \mathbb{R}$.

Proof. As per the hint, we'll show that $\lambda_i(X) \leq \lambda_i(Y)$ for all i. By Courant-Fischer and the given hypotheses, we have

$$\begin{aligned}
0 &\le \lambda_i(Y - X) \\
&= \max_{\dim E = i} \min_{x \in S(E)} \big(\langle Yx, x\rangle - \langle Xx, x\rangle\big) \\
&\le \max_{\dim E = i} \min_{x \in S(E)} \langle Yx, x\rangle - \max_{\dim E = i} \min_{x \in S(E)} \langle Xx, x\rangle \\
&= \lambda_i(Y) - \lambda_i(X).
\end{aligned}$$

The claim follows since the trace is the sum of the eigenvalues. □

(f) Show that $0 \preceq X \preceq Y$ implies $X^{-1} \succeq Y^{-1}$ if X is invertible.

Proof. First, suppose the claim holds when one of the matrices is the identity. We claim that we can multiply $X \preceq Y$ through on the left and right by $Y^{-1/2}$, which gives $Y^{-1/2} X Y^{-1/2} \preceq I$. Since the desired conclusion is assumed to hold when one matrix is the identity, we have that $Y^{1/2} X^{-1} Y^{1/2} \succeq I$. Multiplying through on the right and left by $Y^{1/2}$ gives the desired $X^{-1} \succeq Y^{-1}$.

The product of two psd matrices is psd if and only if the product is also symmetric. Since $Y^{-1/2}(Y - X)Y^{-1/2}$ is symmetric, we have that it is also psd. We then have

$$\begin{aligned}
X \preceq Y &\iff 0 \preceq Y - X \\
&\implies 0 \preceq Y^{-1/2}(Y - X)^{-1/2} \\
&\implies Y^{-1/2} X Y^{-1/2} \preceq I,
\end{aligned}$$

so our earlier claim is justified. It remains to show that the proposition holds for the identity, i.e. $0 \preceq X \preceq I$ implies that $X^{-1} \succeq I$. But this is clear since in part (e) we showed that $0 \preceq X \preceq I$ implies that each eigenvalue of X is between 0 and 1. Consequently, each eigenvalue of X^{-1} is at least 1, so we have $X^{-1} \succeq I$. □

(g) Show that $0 \preceq X \preceq Y$ implies $\log X \preceq \log Y$.

Proof. Since $0 \preceq X \preceq Y$, we have $0 \preceq X + t \preceq Y + t$ for all $t > 0$. By part (f) we deduce that $(X + t)^{-1} \succeq (Y + t)^{-1}$. Finally, by this identity that follows from elementary calculus,

$$\log x = \int_0^\infty \left(\frac{1}{1+t} - \frac{1}{x+t} \right) dt,$$

we have that

$$\log Y - \log X = \int_0^\infty \big((X + t)^{-1} - (Y + t)^{-1} \big)\, dt \succeq 0.$$

□

5.4.11

Let $X_1, \dots, X_N$ be independent, mean zero, $n \times n$ symmetric random matrices, such that $\|X_i\| \le K$ almost surely for all i. Deduce from Bernstein's inequality that

$$E\left\|\sum_{i=1}^N X_i\right\| \lesssim \left\|\sum_{i=1}^N EX_i^2\right\|^{1/2} \sqrt{1+\log n} + K(1+\log n.)$$

Proof. □

5.4.12

Let $\epsilon_1, \dots, \epsilon_n$ be independent symmetric Bernoulli random variables and let $A_1, \dots, A_N$ be symmetric $n \times n$ matrices (deterministic). Prove that for any $t \ge 0$ we have

$$\Pr\left[\left\|\sum_{i=1}^N \epsilon_i A_i\right\| \ge t\right] \le 2n\exp(-t^2/2\sigma^2),$$

where $\sigma^2 = \|\sum_{i=1}^N A_i^2\|$.

Proof. We start with the usual moment generating function setup:

$$\Pr\left[\left\|\sum_{i=1}^N \epsilon_i A_i\right\| \ge t\right] \le e^{-\lambda t} E\exp\left(\lambda\left\|\sum_{i=1}^N \epsilon_i A_i\right\|\right).$$

Now we have that $\|\sum_{i=1}^N \epsilon_i A_i\| = \lambda_1(\sum_{i=1}^N \epsilon_i A_i)$, the magnitude of the largest eigenvalue. As discussed in class, we can bound the maximum eigenvalue of a matrix by its trace, which gives

$$\begin{aligned}
\exp\left(\lambda\left\|\sum_{i=1}^N \epsilon_i A_i\right\|\right) &= \exp\left(\lambda \cdot \lambda_1\left(\sum_{i=1}^N \epsilon_i A_i\right)\right) \\
&= \lambda_1\left[\exp\left(\lambda\sum_{i=1}^N \epsilon_i A_i\right)\right] \\
&\le \operatorname{tr}\exp\left(\lambda\sum_{i=1}^N \epsilon_i A_i\right).
\end{aligned}$$

Now just as in the proof of the matrix Bernstein inequality, we apply Lieb's inequality repeatedly to get

$$E\exp\left(\lambda\left\|\sum_{i=1}^N \epsilon_i A_i\right\|\right) \le \operatorname{tr}\exp\left(\sum_{i=1}^N \log Ee^{\lambda\epsilon_i A_i}\right).$$

We can actually compute this expectation.

$$Ee^{\lambda\epsilon_i A_i} = \cosh(\lambda A_i) \preceq e^{\lambda^2 A_i^2/2}.$$

The semidefinite order bound follows from exercise 5.4.5(b). From 5.4.5(e) we have

$$\operatorname{tr}\exp\left[\sum_{i=1}^N \log E e^{\lambda\epsilon_i A_i}\right] \le \operatorname{tr}\exp\left[\sum_{i=1}^N \log\exp\frac{\lambda^2 A_i^2}{2}\right] = \operatorname{tr}\exp\left(\frac{\lambda^2}{2}\sum_{i=1}^N A_i^2\right).$$

I might be cheating here when I say that $\exp\log A = A$. I think you need some condition on $\|A - I\|$. Now we bound the trace in terms of the maximum eigenvalue.

$$\operatorname{tr}\exp\left(\frac{\lambda^2}{2}\sum_{i=1}^N A_i^2\right) \le n\cdot\lambda_i\left[\exp\left(\frac{\lambda^2}{2}\sum_{i=1}^N A_i^2\right)\right] = n\exp\left(\frac{\lambda^2}{2}\sigma^2\right).$$

Thus, we have

$$\Pr\left[\left\|\sum_{i=1}^N \epsilon_i A_i\right\| \ge t\right] \le n\exp(-\lambda t + \lambda^2\sigma^2/2).$$

The λ that optimizes this bound is $\lambda = t/\sigma^2$, which leaves us with the desired conclusion. □

5.4.13

5.4.15

5.6.8

7.1.8

Consider a random process $(X_t)_{t\in T}$.

(a) Express the increments $\|X_t - X_s\|_2$ [I think this is supposed to be $\|\cdot\|_{L^2}$] in terms of the covariance function $\Sigma(t,s)$.

Solution. We have

$$d(t,s)^2 = E[(X_t - X_s)^2] = E[X_t^2] - 2E[X_tX_s] + E[X_s^2] = \Sigma(t,t) - 2\Sigma(t,s) + \Sigma(s,s).$$

□

(b) Assuming that the zero random variable 0 belongs to the process, express the covariance function $\Sigma(t,s)$ in terms of the increments $\|X_t - X_s\|_{L^2}$.

Solution. Suppose the zero process occurs at $t = t_0$, i.e. $X_{t_0} = 0$ a.s.. We have

$$\begin{aligned}\Sigma(t,s) &= -\frac{1}{2}(-2E[X_tX_s]) \\ &= \frac{1}{2}\big(E[X_t^2] + E[X_s^2] - E[X_t^2 - 2X_tX_s + X_s^2]\big) \\ &= \frac{1}{2}\big(d(t,t_0)^2 + d(s,t_0)^2 - d(t,s)^2\big).\end{aligned}$$

□

7.1.13

Realize an N-step random walk with $Z_i \sim \mathcal{N}(0,1)$ as a canonical Gaussian process with $T \subseteq \mathbb{R}^n$.

Solution. Let t_n be the vector $e_1 + \cdots + e_n$, the sum of the first n standard basis vectors of $\mathbb{R}^N$. If we let $g \sim \mathcal{N}(0, I_N)$, then we have that $\langle g, t_n \rangle$ is a sum of n independent standard normals, which is equal to X_n in distribution. This process is canonical since

$$\|X_n - X_m\|_{L^2} = \|\langle g, t_n - t_m \rangle\|_{L^2} = \sqrt{n-m} = \|t_n - t_m\|_2.$$

$\square$

7.2.4

If $X \sim \mathcal{N}(0, \sigma^2)$, show that

$$EXf(X) = \sigma^2 Ef'(X).$$

Proof. Note that $X = \sigma Z$ for $N \sim \mathcal{N}(0,1)$. We apply Gaussian integration by parts.

$$E[Xf(X)] = \sigma E[Zf(\sigma Z)] = \sigma E[(f(\sigma Z))'] = \sigma^2 E[f(\sigma Z)] = \sigma^2 E[f(X)].$$

$\square$