Financial Engineering and Risk Management

Review of vectors

Martin Haugh **Garud Iyengar**

Columbia University
Industrial Engineering and Operations Research

Reals numbers and vectors

- We will denote the set of real numbers by $\mathbb{R}$
- Vectors are finite collections of real numbers
- Vectors come in two varieties
  - Row vectors: $\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$
  - Column vectors $\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$
  - By default, vectors are column vectors
- The set of all vectors with **n** components is denoted by $\mathbb{R}^{\mathbf{n}}$

Linear independence

- A vector $\mathbf{w}$ is linearly dependent on $\mathbf{v}_1, \mathbf{v}_2$ if

$$\mathbf{w} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 \text{ for some } \alpha_1, \alpha_2 \in \mathbb{R}$$

  Example:

$$\begin{bmatrix} 2 \\ 6 \\ 4 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 4 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

- Other names: linear combination, linear span

- A set $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ are linearly independent if **no** $\mathbf{v}_i$ is linearly dependent on the others, $\{\mathbf{v}_j : j \neq i\}$

Basis

- Every $\mathbf{w} \in \mathbb{R}^n$ is a linear combination of the linearly independent set

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\} \qquad \mathbf{w} = w_1 \underbrace{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{e}_1} + w_2 \underbrace{\begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{e}_2} + \ldots + w_n \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}}_{\mathbf{e}_n}$$

- Basis $\equiv$ any linearly independent set that spans the entire space
- Any basis for $\mathbb{R}^n$ has exactly n elements

Norms

- A function $\rho(\mathbf{v})$ of a vector $\mathbf{v}$ is called a norm if
  - $\rho(\mathbf{v}) \geq 0$ and $\rho(\mathbf{v}) = 0$ implies $\mathbf{v} = \mathbf{0}$
  - $\rho(\alpha\mathbf{v}) = |\alpha|\,\rho(\mathbf{v})$ for all $\alpha \in \mathbb{R}$
  - $\rho(\mathbf{v}_1 + \mathbf{v}_2) \leq \rho(\mathbf{v}_1) + \rho(\mathbf{v}_2)$ (triangle inequality)

  ρ generalizes the notion of "length"

- **Examples**:
  - ℓ_2 norm: $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$ ... usual length
  - ℓ_1 norm: $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$
  - ℓ_∞ norm: $\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x|_i$
  - ℓ_p norm, $1 \leq p < \infty$: $\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x|_i^p\right)^{\frac{1}{p}}$

Inner product

- The inner-product or dot-product of two vector $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ is defined as

$$\mathbf{v} \cdot \mathbf{w} = \sum_{i=1}^{n} v_i w_i$$

- The ℓ_2 norm $\|\mathbf{v}\|_2 = \sqrt{\mathbf{v} \cdot \mathbf{v}}$
- The angle θ between two vectors $\mathbf{v}$ and $\mathbf{w}$ is given by

$$\cos(\theta) = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\|_2 \, \|\mathbf{w}\|_2}$$

- Will show later: $\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^\top \mathbf{w} =$ product of $\mathbf{v}$ transpose and $\mathbf{w}$

Financial Engineering and Risk Management

Review of matrices

Martin Haugh   Garud Iyengar

Columbia University
Industrial Engineering and Operations Research

Matrices

- Matrices are rectangular arrays of real numbers
- Examples:
  - $\mathbf{A} = \begin{bmatrix} 2 & 3 & 7 \\ 1 & 6 & 5 \end{bmatrix}$: 2×3 matrix
  - $\mathbf{B} = \begin{bmatrix} 2 & 3 & 7 \end{bmatrix}$: 1×3 matrix $\equiv$ row vector
  - $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$: $\mathbf{m} \times \mathbf{n}$ matrix ... $\mathbb{R}^{\mathbf{m} \times \mathbf{n}}$
  - $\mathbf{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$ ... $n \times n$ **Identity** matrix
- Vectors are clearly also matrices

Matrix Operations: Transpose

- Transpose: $\mathbf{A} \in \mathbb{R}^{m \times d}$

$$\mathbf{A}^\top = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1d} \\ a_{21} & a_{22} & \dots & a_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{md} \end{bmatrix}^\top = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{d2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1d} & a_{2d} & \dots & a_{md} \end{bmatrix} \in \mathbb{R}^{d \times m}$$

  - Transpose of a row vector is a column vector

- **Example**:
  - $\mathbf{A} = \begin{bmatrix} 2 & 3 & 7 \\ 1 & 6 & 5 \end{bmatrix}$: 2×3 matrix ... $\mathbf{A}^\top = \begin{bmatrix} 2 & 1 \\ 3 & 6 \\ 7 & 5 \end{bmatrix}$: 3×2 matrix
  - $\mathbf{v} = \begin{bmatrix} 2 \\ 6 \\ 4 \end{bmatrix}$: column vector ... $\mathbf{v}^\top = \begin{bmatrix} 2 & 6 & 4 \end{bmatrix}$: row vector

Matrix Operations: Multiplication

- Multiplication: $\mathbf{A} \in \mathbb{R}^{\mathbf{m}\times\mathbf{d}}$, $\mathbf{B} \in \mathbb{R}^{\mathbf{d}\times\mathbf{p}}$ then $\mathbf{C} = \mathbf{AB} \in \mathbb{R}^{\mathbf{m}\times\mathbf{p}}$

$$c_{ij} = \left[\begin{array}{cccc} a_{i1} & a_{i2} & \dots & a_{id} \end{array}\right] \left[\begin{array}{c} b_{1j} \\ b_{2j} \\ \vdots \\ b_{dj} \end{array}\right]$$

  - row vector $\mathbf{v} \in \mathbb{R}^{1\times d}$ times column vector $\mathbf{w} \in \mathbb{R}^{d\times 1}$ is a scalar.
  - Identity times any matrix $\mathbf{AI}_n = \mathbf{I}_m\mathbf{A} = \mathbf{A}$

- **Examples**:
  - $\begin{bmatrix} 2 & 3 & 7 \\ 1 & 6 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 6 \\ 4 \end{bmatrix} = \begin{bmatrix} 2(2) + 3(6) + 7(4) \\ 1(2) + 6(6) + 5(4) \end{bmatrix} = \begin{bmatrix} 50 \\ 58 \end{bmatrix}$
  - ℓ_2 norm: $\left\| \begin{bmatrix} 1 \\ -2 \end{bmatrix} \right\|_2 = \sqrt{1^2 + (-2)^2} = \sqrt{\begin{bmatrix} 1 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix}} = \sqrt{\begin{bmatrix} 1 \\ -2 \end{bmatrix}^\top \begin{bmatrix} 1 \\ -2 \end{bmatrix}}$
  - inner product: $\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^\top \mathbf{w}$

Linear functions

- A function $f : \mathbb{R}^d \mapsto \mathbb{R}^m$ is linear if

$$f(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y}), \quad \alpha, \beta \in \mathbb{R}, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$$

- A function f is linear if and only if $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ for matrix $\mathbf{A} \in \mathbb{R}^{m \times d}$

- **Examples**
  - $f(\mathbf{x}) : \mathbb{R}^3 \mapsto \mathbb{R}$: $f(\mathbf{x}) = \begin{bmatrix} 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 2x_1 + 3x_2 + 4x_3$
  - $f(\mathbf{x}) : \mathbb{R}^3 \mapsto \mathbb{R}^2$: $f(\mathbf{x}) = \begin{bmatrix} 2 & 3 & 4 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_1 + 3x_2 + 4x_3 \\ x_1 + 2x_3 \end{bmatrix}$

- Linear constraints define sets of vectors that satisfy linear relationships
  - Linear equality: $\{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}\}$ ... line, plane, etc.
  - Linear inequality: $\{\mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$ ... half-space

Rank of a matrix

- column rank of $\mathbf{A} \in \mathbb{R}^{m \times d}$ = number of linearly independent columns
  - range($\mathbf{A}$) $= \{\mathbf{y} : \mathbf{y} = \mathbf{A}\mathbf{x}$ for some $\mathbf{x}\}$
  - column rank of $\mathbf{A}$ = size of basis for range($\mathbf{A}$)
  - column rank of $\mathbf{A} = m \Rightarrow$ range($\mathbf{A}$) $= \mathbb{R}^m$

- row rank of $\mathbf{A}$ = number of linearly independent rows
- **Fact**: row rank = column rank $\leq \min\{m, d\}$

- **Example**:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix}, \quad \text{rank} = 1, \quad \text{range}(\mathbf{A}) = \left\{ \lambda \begin{bmatrix} 1 \\ 2 \end{bmatrix} : \lambda \in \mathbb{R} \right\}$$

- $\mathbf{A} \in \mathbb{R}^{\mathbf{n} \times \mathbf{n}}$ and $\mathbf{rank}(\mathbf{A}) = n \Rightarrow \mathbf{A}$ invertible, i.e. $\mathbf{A}^{-1} \in \mathbb{R}^{\mathbf{n} \times \mathbf{n}}$

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

Financial Engineering and Risk Management

Review of linear optimization

Martin Haugh **Garud Iyengar**

Columbia University
Industrial Engineering and Operations Research

Hedging problem

- d assets
- Prices at time $t = 0$: $\mathbf{p} \in \mathbb{R}^d$

- Market in m possible states at time $t = 1$
- Price of asset j in state $i = S_{ij}$

$$\mathbf{S}_j = \begin{bmatrix} S_{1j} \\ S_{2j} \\ \vdots \\ S_{mj} \end{bmatrix} \quad \mathbf{S} = \begin{bmatrix} \mathbf{S}_1 & \mathbf{S}_2 & \dots & \mathbf{S}_d \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1d} \\ S_{21} & S_{22} & \dots & S_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ S_{m1} & S_{m2} & \dots & S_{md} \end{bmatrix} \in \mathbb{R}^{m \times d}$$

- Hedge an obligation $\mathbf{X} \in \mathbb{R}^m$
  - Have to pay X_i if state i occurs
  - Buy/short sell $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^\top$ shares to cover obligation

Hedging problem (contd)

- Position $\boldsymbol{\theta} \in \mathbb{R}^d$ purchased at time $t = 0$
  - θ_j = number of shares of asset j purchased, $j = 1, \ldots, d$
  - Cost of the position $\boldsymbol{\theta} = \sum_{j=1}^d p_j \theta_j = \mathbf{p}^\top \boldsymbol{\theta}$

- Payoff from liquidating position at time $t = 1$
  - payoff y_i in state i: $y_i = \sum_{j=1}^d S_{ij}\theta_j$
  - Stacking payoffs for all states: $\mathbf{y} = \mathbf{S}\boldsymbol{\theta}$
  - Viewing the payoff vector $\mathbf{y}$: $\mathbf{y} \in \text{range}(\mathbf{S})$

$$\mathbf{y} = \begin{bmatrix} \mathbf{S}_1 & \mathbf{S}_2 & \ldots & \mathbf{S}_d \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{bmatrix} = \sum_{j=1}^d \theta_j \mathbf{S}_j$$

- Payoff $\mathbf{y}$ hedges $\mathbf{X}$ if $\mathbf{y} \geq \mathbf{X}$.

Hedging problem (contd)

- Optimization problem:

$$
\begin{array}{rl}
\text{min} & \sum_{j=1}^{d} p_j\theta_j \quad (\equiv \mathbf{p}^\top \boldsymbol{\theta}) \\
\text{subject to} & \sum_{j=1}^{d} S_{ij}\theta_j \geq X_i, \quad i = 1, \ldots, m \quad (\equiv \mathbf{S}\boldsymbol{\theta} \geq \mathbf{X})
\end{array}
$$

- Features of this optimization problem
  - Linear objective function: $\mathbf{p}^\top \boldsymbol{\theta}$
  - Linear inequality constraints: $\mathbf{S}\boldsymbol{\theta} \geq \mathbf{X}$

- Example of a linear program
  - Linear objective function: either a min/max
  - Linear inequality and equality constraints

$$
\begin{array}{rl}
\text{max/min}_{\mathbf{x}} & \mathbf{c}^\top \mathbf{x} \\
\text{subject to} & \mathbf{A}_{eq}\mathbf{x} = \mathbf{b}_{eq} \\
 & \mathbf{A}_{in}\mathbf{x} \leq \mathbf{b}_{in}
\end{array}
$$

Linear programming duality

- Linear program

$$P = \begin{array}{rl} \min_{\mathbf{x}} & \mathbf{c}^\top \mathbf{x} \\ \text{subject to} & \mathbf{A}\mathbf{x} \geq \mathbf{b} \end{array}$$

- Dual linear program

$$D = \begin{array}{rl} \max_{\mathbf{u}} & \mathbf{b}^\top \mathbf{u} \\ \text{subject to} & \mathbf{A}^\top \mathbf{u} = \mathbf{c} \\ & \mathbf{u} \geq \mathbf{0} \end{array}$$

Theorem.

- **Weak Duality**: $P \geq D$
- **Bound**: $\mathbf{x}$ feasible for P, $\mathbf{u}$ feasible for D, $\mathbf{c}^\top \mathbf{x} \geq P \geq D \geq \mathbf{b}^\top \mathbf{u}$
- **Strong Duality**: Suppose P or D finite. Then $P = D$.
- Dual of the dual is the primal (original) problem

More duality results

- Here is another primal-dual pair

$$\begin{array}{rl} \min_{\mathbf{x}} & \mathbf{c}^\top \mathbf{x} \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \end{array} \quad = \quad \begin{array}{rl} \max_{\mathbf{u}} & \mathbf{b}^\top \mathbf{u} \\ \text{subject to} & \mathbf{A}^\top \mathbf{u} = \mathbf{c} \end{array}$$

- General idea for constructing duals

$$\begin{aligned} P &= \min\{\mathbf{c}^\top \mathbf{x} : \mathbf{A}\mathbf{x} \geq \mathbf{b}\} \\ &\geq \min\{\mathbf{c}^\top \mathbf{x} - \mathbf{u}^\top (\mathbf{A}\mathbf{x} - \mathbf{b}) : \mathbf{A}\mathbf{x} \geq \mathbf{b}\} \text{ for all } \mathbf{u} \geq \mathbf{0} \\ &\geq \mathbf{b}^\top \mathbf{u} + \min\{(\mathbf{c} - \mathbf{A}^\top \mathbf{u})^\top \mathbf{x} : \mathbf{x} \in \mathbb{R}^n\} \\ &= \begin{cases} \mathbf{b}^\top \mathbf{u} & \mathbf{A}^\top \mathbf{u} = \mathbf{c} \\ -\infty & \text{otherwise} \end{cases} \\ &\geq \max\{\mathbf{b}^\top \mathbf{u} : \mathbf{A}^\top \mathbf{u} = \mathbf{c}\} \end{aligned}$$

- Lagrangian relaxation: dualize constraints and relax them!

Financial Engineering and Risk Management

Review of nonlinear optimization

Martin Haugh **Garud Iyengar**

Columbia University
Industrial Engineering and Operations Research

Unconstrained nonlinear optimization

- Optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$$

- Categorization of minimum points
  - $\mathbf{x}^*$ global minimum if $f(\mathbf{y}) \geq f(\mathbf{x}^*)$ for all $\mathbf{y}$
  - $\mathbf{x}^*_{loc}$ local minimum if $f(\mathbf{y}) \geq f(\mathbf{x}^*_{loc})$ for all $\mathbf{y}$ such that $\|\mathbf{y} - \mathbf{x}^*_{loc}\| \leq r$

- Sufficient condition for local min
  - gradient $\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \mathbf{0}$: local stationarity
  - Hessian $\nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$ positive semidefinite

- Gradient condition is sufficient if the function $f(\mathbf{x})$ is convex.

Unconstrained nonlinear optimization

- Optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^2}\ x_1^2 + 3x_1x_2 + x_2^3$$

- Gradient

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2x_1 + 3x_2 \\ 3x_1 + 3x_2^2 \end{bmatrix} = \mathbf{0} \quad \Rightarrow \quad \mathbf{x} = \mathbf{0}, \quad \begin{bmatrix} -\frac{9}{4} \\ \frac{3}{2} \end{bmatrix}$$

- Hessian at $\mathbf{x}$: $\mathbf{H} = \begin{bmatrix} 2 & 3 \\ 3 & 6x_2 \end{bmatrix}$
    - $\mathbf{x} = \mathbf{0}$: $\mathbf{H} = \begin{bmatrix} 2 & 3 \\ 3 & 0 \end{bmatrix}$. Not positive definite. Not local minimum.
    - $\mathbf{x} = \begin{bmatrix} -\frac{9}{4} \\ \frac{3}{2} \end{bmatrix}$: $\mathbf{H} = \begin{bmatrix} 2 & 3 \\ 3 & 9 \end{bmatrix}$. Positive semidefinite. Local minimum

Lagrangian method

- Constrained optimization problem

$$\begin{aligned} \max_{\mathbf{x}\in\mathbb{R}^2} \quad & 2\ln(1+x_1) + 4\ln(1+x_2), \\ \text{s.t.} \quad & x_1 + x_2 = 12 \end{aligned}$$

- Convex problem. But constraints make the problem hard to solve.
- Form a Lagrangian function

$$\mathcal{L}(\mathbf{x}, v) = 2\ln(1+x_1) + 4\ln(1+x_2) - v(x_1 + x_2 - 12)$$

- Compute the stationary points of the Lagrangian as a function of v

$$\nabla\mathcal{L}(\mathbf{x}, v) = \begin{bmatrix} \frac{2}{1+x_1} - v \\ \frac{4}{1+x_2} - v \end{bmatrix} = \mathbf{0} \quad \Rightarrow \quad x_1 = \frac{2}{v} - 1, \quad x_2 = \frac{4}{v} - 1$$

- Substituting in the constraint $x_1 + x_2 = 12$, we get

$$\frac{6}{v} = 14 \quad \Rightarrow \quad v = \frac{3}{7} \quad \Rightarrow \quad \mathbf{x} = \frac{1}{3}\begin{bmatrix} 11 \\ 25 \end{bmatrix}$$

Portfolio Selection

- Optimization problem

$$\begin{aligned} \max_{\mathbf{x}} \quad & \boldsymbol{\mu}^\top \mathbf{x} - \lambda \mathbf{x}^\top \mathbf{V} \mathbf{x} \\ \text{s.t.} \quad & \mathbf{1}^\top \mathbf{x} = 1 \end{aligned}$$

  Constraints make the problem hard!

- Lagrangian function

$$\mathcal{L}(\mathbf{x}, v) = \boldsymbol{\mu}^\top \mathbf{x} - \lambda \mathbf{x}^\top \mathbf{V} \mathbf{x} - v(\mathbf{1}^\top \mathbf{x} - 1)$$

  - Solve for the maximum value with no constraints

$$\boldsymbol{\nabla}_x \mathcal{L}(\mathbf{x}, v) = \boldsymbol{\mu} - 2\lambda \mathbf{V}\mathbf{x} - v\mathbf{1} = \mathbf{0} \quad \Rightarrow \quad \mathbf{x} = \frac{1}{2\lambda} \cdot \mathbf{V}^{-1}(\boldsymbol{\mu} - v\mathbf{1})$$

  - Solve for v from the constraint

$$\mathbf{1}^\top \mathbf{x} = 1 \quad \Rightarrow \quad \mathbf{1}^\top \mathbf{V}^{-1}(\boldsymbol{\mu} - v\mathbf{1}) = 2\lambda \quad \Rightarrow \quad v = \frac{\mathbf{1}^\top \mathbf{V}^{-1} \boldsymbol{\mu} - 2\lambda}{\mathbf{1}^\top \mathbf{V}^{-1} \mathbf{1}}$$

  - Substitute back in the expression for $\mathbf{x}$