Portfolio Selection and Convex Optimization

2019 Winter Camp at PHBS

Xianhua Peng
Associate Professor
PHBS, Peking University

January 2019

Outline

Outline

Portfolio return

- there are n risky assets that we can invest in
- $\tilde{p}_i(t)$: the price of the ith asset at time t (tth second, mins, hours, days, weeks, months)
- random simple return from $t-1$ to t: $\tilde{r}_i(t) = \frac{\tilde{p}_i(t) - p_i(t-1)}{p_i(t-1)}$, $i = 1, \dots, n$
- mean and covariance matrix of $\tilde{r}(t) = (\tilde{r}_1(t), \dots, \tilde{r}_n(t))'$:

$$\mu = E[\tilde{r}(t)]$$
$$V(t) = Cov(\tilde{r}(t)) = E[(\tilde{r}(t) - \mu)(\tilde{r}(t) - \mu)']$$

- $V(t)$ is positive semidefinite $\Leftrightarrow$ all eigenvalues ≥ 0
- $V(t)$ is invertible $\Leftrightarrow$ all eigenvalues are strictly positive
- $V(t)$ has a zero eigenvalue $\Leftrightarrow$ at least one asset return is a linear function of the other asset returns

Portfolio return (cont'd)

- Diagonal elements of $V(t)$:
  $V_{ii}(t) = cov(\tilde{r}_i(t), \tilde{r}_i(t)) = var(\tilde{r}_i(t)) = \sigma_i^2$
- Off-diagonal elements of $V(t)$: $V_{ij}(t) = cov(\tilde{r}_i(t), \tilde{r}_j(t)) = \rho_{ij}\sigma_i\sigma_j$
- Correlation coefficient $\rho_{ij} \in [-1, 1]$
  - $\rho_{ij} = -1$: perfectly negatively correlated
  - $\rho_{ij} = 1$: perfectly positively correlated
- Decomposition of the variance-covariance matrix

$$V = \mathrm{diag}(\sigma_1, \ldots, \sigma_n) \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{bmatrix} \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$$

Portfolio return (cont'd)

- at time $t-1$, we have initial wealth \$$v_0$. We construct a portfolio with a position $x = (x_1, \ldots, x_n)'$ at time $t-1$, x_i is the amount invested in asset i, $i = 1, \ldots, n$, $x_i > 0$ means long position in asset i, $x_i < 0$ means short position in asset i
- portfolio weight vector: $\alpha = (\alpha_1, \ldots, \alpha_n)'$: $\alpha_i := \frac{x_i}{v_0}$ is the fraction invested in asset i, α_i can be positive or negative
- $\sum_{i=1}^n x_i = v_0$, $\sum_{i=1}^n \alpha_i = 1$

Portfolio return (cont'd)

- simple return of the portfolio:
  $\tilde{r}_x(t) = \frac{\sum_{i=1}^n \tilde{r}_i(t)x_i}{v_0} = \sum_{i=1}^n \tilde{r}_i(t)\alpha_i = \alpha'\tilde{r}(t)$
- The return $\tilde{r}_x(t)$ only depends on the weight α, so we denote it also as $\tilde{r}_\alpha(t)$
- Expected portfolio return: $E[\tilde{r}_\alpha(t)] = \alpha' E[\tilde{r}(t)] = \alpha'\mu$
- variance of portfolio return:

$$\begin{aligned} var(\tilde{r}_\alpha(t)) &= E[(\tilde{r}_\alpha(t) - E[\tilde{r}_\alpha(t)])(\tilde{r}_\alpha(t) - E[\tilde{r}_\alpha(t)])'] \\ &= E[\alpha'(\tilde{r}(t) - E[\tilde{r}(t)])(\alpha'(\tilde{r}(t) - E[\tilde{r}(t)]))'] \\ &= E[\alpha'(\tilde{r}(t) - \mu)(\tilde{r}(t) - \mu)'\alpha] \\ &= \alpha' E[(\tilde{r}(t) - \mu)(\tilde{r}(t) - \mu)']\alpha \\ &= \alpha' V(t)\alpha \end{aligned}$$

Maximizing expected utility

In Economics, one models investor's preference over a random payoff W using expected utility $E[u(W)]$:

- $u'(w) \geq 0$: u is non-decreasing
- $u''(w) \leq 0$: u is concave, as investor is risk averse
- Utility maximization problem:

$$\max_{\alpha} E[u(v_0 \cdot (1 + \alpha' \tilde{r}))]$$
$$\text{s.t. } \mathbf{1}'\alpha = 1$$

- typical utility functions
  - logarithmic utility $u(x) = \log(x)$
  - exponential utility $u(x) = -e^{-\nu x}$, $\nu > 0$
  - power utility $u(x) = \dfrac{x^{1-\alpha}}{1-\alpha}, \alpha > 0$

Mean-variance optimization problem

Suppose one of the following two conditions hold:
(i) $\tilde{r}$ has a multivariate normal distribution, or
(ii) utility function $u(x) = ax - bx^2$, $b > 0$.
Suppose an investor wants to have an expected return $E[\tilde{r}_\alpha] = r$.
Then, the utility maximization problem

$$\begin{aligned} \max_{\alpha} \; & E[u(v_0 \cdot (1 + \tilde{r}_\alpha))] \\ \text{s.t. } & E[\tilde{r}_\alpha] = r \\ & \mathbf{1}'\alpha = 1 \end{aligned}$$

is equivalent to

$$\begin{aligned} \min_{\alpha} \; & var(v_0 \cdot (1 + \alpha'\tilde{r})) = v_0^2 \alpha' V \alpha \\ \text{s.t. } & \mu'\alpha = r \\ & \mathbf{1}'\alpha = 1. \end{aligned}$$

Solving the mean-variance problem

$$\begin{aligned}\min_{\alpha}\ & \frac{1}{2}\alpha' V \alpha \\ \text{s.t. } & \mu'\alpha = r \\ & \mathbf{1}'\alpha = 1.\end{aligned}$$

- Lagrangian function

$$\mathcal{L}(\alpha, \nu_1, \nu_2) = \frac{1}{2}\alpha' V \alpha + \nu_1(\mu'\alpha - r) + \nu_2(\mathbf{1}'\alpha - 1)$$

- Take the gradient with respect to α and set it zero:

$$V\alpha + \nu_1 \mu + \nu_2 \mathbf{1} = 0 \Rightarrow \alpha = -\nu_1 V^{-1}\mu - \nu_2 V^{-1}\mathbf{1}.$$

Solving the mean-variance problem (cont'd)

- Substituting this back into the constraints we get

$$-(\mu' V^{-1}\mu)\nu_1 - (\mu' V^{-1}\mathbf{1})\nu_2 = r$$

$$-(\mathbf{1}' V^{-1}\mu)\nu_1 - (\mathbf{1}' V^{-1}\mathbf{1})\nu_2 = 1$$

  Hence,

$$\begin{bmatrix}\nu_1\\ \nu_2\end{bmatrix} = M^{-1}\begin{bmatrix}r\\ 1\end{bmatrix}, \text{ where } M = \begin{bmatrix}-\mu' V^{-1}\mu & -\mu' V^{-1}\mathbf{1}\\ -\mathbf{1}' V^{-1}\mu & -\mathbf{1}' V^{-1}\mathbf{1}\end{bmatrix}$$

- the optimal α is

$$\alpha = [-V^{-1}\mu, -V^{-1}\mathbf{1}]\begin{bmatrix}\nu_1\\ \nu_2\end{bmatrix} = [-V^{-1}\mu, -V^{-1}\mathbf{1}]M^{-1}\begin{bmatrix}r\\ 1\end{bmatrix} = \gamma_1 r + \gamma_0$$

Efficient frontier

- Efficient frontier:

$$\begin{aligned}\sigma^2(r) &= \alpha' V \alpha \\ &= (r\gamma_1 + \gamma_0)' V (r\gamma_1 + \gamma_0) \\ &= (\gamma_1' V \gamma_1) r^2 + 2\gamma_0' V \gamma_1 r + \gamma_0' V \gamma_0\end{aligned}$$

- Minimum variance portfolio

$$\begin{aligned}\sigma^2_{\min} = \min_{\alpha} & \ \alpha' V \alpha \\ \text{s.t. } & \mathbf{1}'\alpha = 1.\end{aligned}$$

Mean-variance with risk-free asset

- $n+1$ assets: n risky assets with returns $\tilde{r}$ and a risk-free asset with return r_f
- α: fraction invested in risky assets
- α_0: fraction invested in the risk-free asset
- Return of the portfolio (α_0, α): $\alpha'\tilde{r} + \alpha_0 r_f$. Mean return is $\mu'\alpha + \alpha_0 r_f$, where $\mu = E[\tilde{r}]$. Variance of return is $\alpha' V\alpha$.
- Mean-variance portfolio selection with risk-free asset

$$
\begin{aligned}
\min_{\alpha} \ & \frac{1}{2}\alpha' V\alpha \\
\text{s.t. } & \alpha_0 r_f + \mu'\alpha = r \\
& \alpha_0 + \mathbf{1}'\alpha = 1.
\end{aligned}
$$

- Solve for α_0: $\alpha_0 = 1 - \mathbf{1}'\alpha$
- Substitute in the other constrain: $(\mu - r_f\mathbf{1})'\alpha = r - r_f$
- $\hat{r} = r - r_f$ is the excess target return, $\hat{\mu} = \mu - r_f\mathbf{1}$ is the excess return

Mean-variance with risk-free asset (cont'd)

- Equivalent problem

$$\min_{\alpha} \frac{1}{2}\alpha' V \alpha$$
$$\text{s.t. } \hat{\mu}'\alpha = \hat{r}$$

- Lagrangian function: $\mathcal{L} = \frac{1}{2}\alpha' V \alpha + \nu(\hat{\mu}'\alpha - \hat{r})$
- Set gradient to 0: $\nabla_\alpha \mathcal{L} = V\alpha + \nu\hat{\mu} = 0 \Rightarrow \alpha = -\nu V^{-1}\hat{\mu}$
- Substitute into constraint: $\hat{r} = \hat{\mu}'\alpha = -\nu\hat{\mu}' V^{-1}\hat{\mu} \Rightarrow \nu = -\dfrac{\hat{r}}{\hat{\mu}' V^{-1}\hat{\mu}}$
- Thus, $\alpha^* = \hat{r} \cdot \frac{V^{-1}\hat{\mu}}{\hat{\mu}' V^{-1}\hat{\mu}} = \hat{r}\gamma^*$
- Efficient frontier is a straight line:

$$\sigma(r) = \sqrt{(\alpha^*)' V \alpha^*} = \hat{r} \cdot \frac{1}{\sqrt{\hat{\mu}' V^{-1}\hat{\mu}}}$$

Outline

No-short sales constraints

$$\begin{aligned} \min_{\alpha} \ & \frac{1}{2}\alpha' V \alpha \\ \text{s.t. } & \mu'\alpha = r \\ & \mathbf{1}'\alpha = 1 \\ & \alpha \geq 0. \end{aligned}$$

- No closed form solution available.
- The efficient numerical algorithm is the interior-point algorithm.

Margin constraints

- $\alpha_i^+ = \max\{\alpha_i, 0\}$ is the long position in asset i
- $\alpha_i^- = \max\{-\alpha_i, 0\}$ is the short position in asset i
- $\alpha_i = \alpha_i^+ - \alpha_i^-$ is the (net) position in asset i
- Margin constraint $M \sum_{i=1}^n \alpha_i^- \leq \sum_{i=1}^n \alpha_i^+$. Since $\sum_{i=1}^n \alpha_i = \sum_{i=1}^n (\alpha_i^+ - \alpha_i^-) = 1$, the margin constraint is equivalent to $\sum_{i=1}^n \alpha_i^- \leq \frac{1}{M-1}$.
- optimization problem

$$
\begin{aligned}
\min_{\alpha}\ & \frac{1}{2}\alpha' V \alpha \\
\text{s.t. } & \sum_{i=1}^n \alpha_i^- \leq \frac{1}{M-1} \\
& \alpha_i^- + \alpha_i \geq 0,\ i = 1, \ldots, n \\
& \mu' \alpha = r \\
& \mathbf{1}' \alpha = 1 \\
& \alpha_i^- \geq 0, i = 1, \ldots, n
\end{aligned}
$$

Margin constraints (cont'd)

- Then equivalent optimization problem

$$
\begin{aligned}
\min_{\alpha}\ & \begin{bmatrix} \alpha \\ \alpha^- \end{bmatrix}' \begin{bmatrix} V & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \alpha^- \end{bmatrix} \\
\text{s.t. } & \begin{bmatrix} 0 & 1' \\ -I & -I \end{bmatrix} \begin{bmatrix} \alpha \\ \alpha^- \end{bmatrix} \leq \begin{bmatrix} \frac{1}{M-1} \\ 0 \end{bmatrix} \\
& \begin{bmatrix} \mu' & 0 \\ 1' & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \alpha^- \end{bmatrix} = \begin{bmatrix} r \\ 1 \end{bmatrix} \\
& \begin{bmatrix} \alpha \\ \alpha^- \end{bmatrix} \geq \begin{bmatrix} -\infty 1 \\ 0 \end{bmatrix}
\end{aligned}
$$

Diversification constraints

- The mean-variance model tends to produce portfolios with unreasonably large long or short positions in certain asset classes. This issue is well documented in the literature and is often attributed to estimation errors.
- Hence, portfolios chosen by this mean-variance program may be subject to idiosyncratic risk.
- Practitioners often use additional diversification constraints to ensure that the chosen portfolio is well diversified: $\alpha_i \leq m, i = 1, \ldots, m$ or $|\alpha_i| \leq m, i = 1, \ldots, m$
- One can also reduce sector risk by grouping together investments in securities of a sector and setting a limit on the exposure to this sector. If m_k is the maximum that can be invested in sector k, we add the constraint

$$\sum_{i \text{ in sector } k} \alpha_i \leq m_k.$$

Turnover constraints

- We can add a portfolio turnover constraint to ensure that the change between the current holdings α^0 and the desired portfolio α is bounded by h.
- This constraint is essential when solving large mean-variance models since the covariance matrix is almost singular in most practical applications and hence the optimal decision can change significantly with small changes in the problem data. To avoid big changes when reoptimizing the portfolio, turnover constraints are imposed.
- Let y_i be the amount of asset i bought and z_i the amount sold. The turnover constraint is

$$\alpha_i - \alpha_i^0 \leq y_i, y_i \geq 0, i = 1, \ldots, n$$

$$\alpha_i^0 - \alpha_i \leq z_i, z_i \geq 0, i = 1, \ldots, n$$

$$\sum_{i=1}^{n} (y_i + z_i) \leq h$$

Transaction cost constraint

- Suppose that there is a transaction cost t_i (t_i') proportional to the amount of asset i bought (sold).
- Suppose the portfolio is reoptimized once per period. Let α^0 denote the current portfolio. Then a reoptimized portfolio is obtained by solving

$$
\begin{aligned}
\min_{\alpha}\ & \frac{1}{2}\alpha' V \alpha \\
\text{s.t.}\ & \sum_{i=1}^{n}(\mu_i \alpha_i - t_i y_i - t_i' z_i) \geq r \\
& \mathbf{1}'\alpha = 1 \\
& \alpha_i - \alpha_i^0 \leq y_i, y_i \geq 0, i = 1, \ldots, n \\
& \alpha_i^0 - \alpha_i \leq z_i, z_i \geq 0, i = 1, \ldots, n
\end{aligned}
$$

β constraint

- We may add constraints on the upper or lower bound of the β of the portfolio
- Suppose the beta of the ith asset is β_i, $i = 1, \ldots, n$.
- The β constraint is

$$\sum_{i=1}^{n} \beta_i \alpha_i \leq u$$

Outline

Optimization problem in standard form

$$\begin{aligned} \text{minimize} \ \ & f_0(x), \\ \text{subject to } & f_i(x) \leq 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p, \end{aligned}$$

- $x \in \mathrm{R}^n$ is the optimization variable
- $f_0 : \mathrm{R}^n \to \mathrm{R}$ is the objective or cost function
- $f_i : \mathrm{R}^n \to \mathrm{R}$ are the inequality constraint functions
- $h_i : \mathrm{R}^n \to \mathrm{R}$ are the equality constraint functions

optimal value is denoted as $p^\star$

- $p^\star = \infty$ if problem is infeasible (no x satisfies the constraints)
- $p^\star = -\infty$ if problem is unbounded below

Implicit constraints

the standard form optimization problem has an implicit constraint

$$x \in \mathcal{D} := \bigcap_{i=0}^{m} \mathrm{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathrm{dom}\, h_i$$

- we call $\mathcal{D}$ the domain of the problem
- the constraints $f_i(x) \leq 0, h_i(x) = 0$ are the explicit constraints
- a problem is unconstrained if it has no explicit constraints $(m = p = 0)$

example: minimize $f_0(x) = -\sum_{i=1}^{k} \log(b_i - a_i^T x)$ is an unconstrained problem with implicit constraints $a_i^T x < b_i$

Optimal and locally optimal points

x is feasible if $x \in \mathcal{D}$ and it satisfies the explicit constraints

a feasible x is optimal if $f_0(x) = p^\star$; X_{opt} is the set of optimal points

x is locally optimal if there exists an $R > 0$ such that x is optimal for

$$\begin{aligned} &\text{minimize (over } z) \; f_0(z), \\ &\quad \text{subject to } f_i(z) \leq 0, i = 1, \ldots, m \\ &\quad\quad h_i(z) = 0, i = 1, \ldots, p, \\ &\quad\quad \|z - x\|_2 \leq R \end{aligned}$$

examples (with $n = 1, m = p = 0$)

- $f_0(x) = x \log x$, $\mathrm{dom}\, f_0 = \mathrm{R}_{++}$, $p^\star = -1/e$, $x = 1/e$ is optimal
- $f_0(x) = x^3 - 3x$, $p^\star = -\infty$, local optimum at $x = 1$

Convex functions

- $f : \mathrm{R}^n \to \mathrm{R}$ is convex if $\mathrm{dom}\, f$ is a convex set and

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$$

  for all $x, y \in \mathrm{dom}\, f, 0 \leq \theta \leq 1$
- f is concave if $-f$ is convex
- f is strictly convex if $\mathrm{dom}\, f$ is convex and

$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta)f(y)$$

  for $x, y \in \mathrm{dom}\, f, 0 < \theta < 1$

Convex functions - first-order condition

- f is differentiable if $\mathrm{dom}\, f$ is open and the gradient

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n} \right)$$

  exists at each $x \in \mathrm{dom}\, f$
- first-order condition: differentiable f with convex domain is convex iff

$$f(y) \geq f(x) + \nabla f(x)^T (y - x), \text{for all } x, y \in \mathrm{dom}\, f$$

- first-order approximation of convex f is global underestimator

Convex functions - second-order condition

- f is twice differentiable if $\mathrm{dom}\, f$ is open and the Hessian $\nabla^2 f(x)$:

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \; i,\, j = 1, \ldots, n,$$

  exists at each $x \in \mathrm{dom}\, f$
- *2nd*-order conditions: for twice differentiable f with convex domain, f is convex if and only if

$$\nabla^2 f(x) \text{ is positive semi-definite for all } x \in \mathrm{dom}\, f$$

- if $\nabla^2 f(x)$ is positive definite for all $x \in \mathrm{dom}\, f$, then f is strictly convex

Convex optimization problems

standard form convex optimization problem

$$
\begin{aligned}
\text{minimize} \;\; & f_0(x), \\
\text{subject to} \; & f_i(x) \leq 0, i = 1, \ldots, m \\
& a_i^T x = b_i, i = 1, \ldots, p,
\end{aligned}
$$

where $f_0, f_1, \ldots, f_m$ are convex; equality constraints are affine.
often written as

$$
\begin{aligned}
\text{minimize} \;\; & f_0(x), \\
\text{subject to} \; & f_i(x) \leq 0, i = 1, \ldots, m \\
& Ax = b.
\end{aligned}
$$

important property: feasible set of a convex optimization problem is convex

Local and global optima

Any locally optimal point of a convex problem is (globally) optimal.
proof: Suppose x is locally optimal, but there exists a feasible y with $f_0(y) < f_0(x)$. x is locally optimal means there is an $R > 0$ such that:
z is feasible, $\|z - x\|_2 \leq R \Rightarrow f_0(z) \geq f_0(x)$.
Consider $z = \theta y + (1 - \theta)x$ with $\theta = R/(2\|y - x\|_2)$. Then,
1. $\|y - x\|_2 > R$, so $0 < \theta < 1/2$
2. z is a convex combination of two feasible points, hence also feasible
3. $\|z - x\|_2 = R/2$ and

$$f_0(z) \leq \theta f_0(y) + (1 - \theta)f_0(x) < f_0(x),$$

which contradicts our assumption that x is locally optimal

Lagrangian

Standard form problem (not necessarily convex)

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p, \end{aligned}$$

where variable $x \in \mathrm{R}^n$, domain $\mathcal{D}$, optimal value $p^\star$.
Lagrangian: $L : \mathrm{R}^n \times \mathrm{R}^m \times \mathrm{R}^p \to \mathrm{R}$, with $\mathrm{dom}\, L = \mathcal{D} \times \mathrm{R}^m \times \mathrm{R}^p$,

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $f_i(x) \leq 0$
- ν_i is Lagrange multiplier associated with $h_i(x) = 0$

Lagrange dual function

Lagrange dual function: $g : \mathrm{R}^m \times \mathrm{R}^p \rightarrow \mathrm{R}$,

$$g(\lambda, \nu) = \inf_{x \in D} L(x, \lambda, \nu)$$

$$= \inf_{x \in D} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

- g is concave, can be $-\infty$ for some λ, ν
- lower bound property: if $\lambda \geq 0$, then $g(\lambda, \nu) \leq p^\star$
  proof: if $\tilde{x}$ is feasible and $\lambda \geq 0$, then

$$f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \inf_{x \in D} L(x, \lambda, \nu) = g(\lambda, \nu)$$

  minimizing over all feasible $\tilde{x}$ gives $p^\star \geq g(\lambda, \nu)$

Example: Least-norm solution of linear equations

$$\begin{aligned} &\text{minimize} \;\; x^T x \\ &\text{subject to } Ax = b \end{aligned}$$

- Lagrangian is $L(x, \nu) = x^T x + \nu^T (Ax - b)$
- To minimize L over x, set gradient equal to zero:

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0 \;\Rightarrow\; x = -(1/2)A^T \nu$$

- Plug in L to obtain the dual function g:

$$g(\nu) = L((-1/2)A^T \nu, \nu) = -\frac{1}{4}\nu^T A A^T \nu - b^T \nu$$

- g is a concave function of ν
- lower bound property: $p^\star \geq -(1/4)\nu^T A A^T \nu - b^T \nu$ for all ν

Example: Standard form LP

$$
\begin{aligned}
\text{minimize} \ \ & c^T x \\
\text{subject to} \ & Ax = b \\
& x \geq 0
\end{aligned}
$$

- Lagrangian is

$$L(x,\ \lambda,\ \nu) = c^T x + \nu^T (Ax - b) - \lambda^T x = -b^T \nu + (c + A^T \nu - \lambda)^T x$$

- L is affine in x, hence

$$g(\lambda,\ \nu) = \inf_x\ L(x,\ \lambda,\ \nu) = \begin{cases} -b^T \nu & \text{if } A^T \nu - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

- g is linear on affine domain $\{(\lambda,\ \nu) | A^T \nu - \lambda + c = 0, \lambda \geq 0\}$, hence concave
- lower bound property: $p^\star \geq -b^T \nu$ if $A^T \nu + c \geq 0$

The dual problem

Lagrange dual problem

$$\begin{aligned} \text{maximize} \quad & g(\lambda, \nu) \\ \text{subject to} \quad & \lambda \geq 0 \end{aligned}$$

- finds best lower bound on $p^\star$, obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted $d^\star$
- λ, ν are dual feasible if $\lambda \geq 0, (\lambda, \nu) \in \operatorname{dom} g$
- often simplified by making implicit constraint $(\lambda, \nu) \in \operatorname{dom} g$ explicit

example: standard form LP and its dual

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned} \qquad \begin{aligned} \text{maximize} \quad & -b^T \nu \\ \text{subject to} \quad & A^T \nu + c \geq 0 \end{aligned}$$

Weak and strong duality

weak duality: $d^\star \leq p^\star$

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems

strong duality: $d^\star = p^\star$

- does not hold in general
- (usually) holds for convex problems
- conditions that guarantee strong duality in convex problems are called **constraint qualifications**

Slater's constraint qualification

strong duality holds for a convex problem

$$\begin{aligned} \text{minimize} \;\; & f_0(x) \\ \text{subject to} \; & f_i(x) \leq 0, i = 1, \ldots, m \\ & Ax = b, \end{aligned}$$

if it is strictly feasible, *i.e.*,

$$\exists x \in \operatorname{int} \mathcal{D} : \; f_i(x) < 0, i = 1, \ldots, m, \; Ax = b$$

- also guarantees that the dual optimum is attained (if $p^\star > -\infty$)
- can be sharpened: *e.g.*, can replace $\operatorname{int} \mathcal{D}$ with $\operatorname{relint} \mathcal{D}$ (interior relative to affine hull); linear inequalities do not need to hold with strict inequality, ...
- there exist many other types of constraint qualifications

Inequality form LP

primal problem

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax \leq b \end{aligned}$$

dual function

$$g(\lambda) = \inf_x \left((c + A^T\lambda)^T x - b^T\lambda\right) = \begin{cases} -b^T\lambda & \text{if } A^T\lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

dual problem

$$\begin{aligned} \text{maximize} \quad & -b^T\lambda \\ \text{subject to} \quad & A^T\lambda + c = 0 \\ & \lambda \geq 0 \end{aligned}$$

- from Slater's condition: $p^\star = d^\star$ if $A\tilde{x} < b$ for some $\tilde{x}$
- in fact, $p^\star = d^\star$ except when primal or dual are infeasible

Quadratic program

primal problem (assume $P \in \mathbf{S}^n_{++}$)

$$\begin{array}{ll} \text{minimize} & x^T P x \\ \text{subject to} & Ax \leq b \end{array}$$

dual function

$$g(\lambda) = \inf_x \left(x^T P x + \lambda^T (Ax - b)\right) = -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

dual problem

$$\begin{array}{ll} \text{maximize} & -(1/4)\lambda^T A P^{-1} A^T \lambda - b^T \lambda \\ \text{subject to} & \lambda \geq 0 \end{array}$$

- from Slater's condition: $p^\star = d^\star$ if $A\tilde{x} < b$ for some $\tilde{x}$
- in fact, $p^\star = d^\star$ always

Complementary slackness

assume strong duality holds, $x^\star$ is primal optimal, $(\lambda^\star, \nu^\star)$ is dual optimal

$$
\begin{aligned}
f_0(x^\star) = g(\lambda^\star, \nu^\star) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\
&\leq f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \\
&\leq f_0(x^\star)
\end{aligned}
$$

hence, the two inequalities hold with equality

- $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$
- $\lambda_i^\star f_i(x^\star) = 0$ for $i = 1, \ldots, m$ (known as complementary slackness):

$$
\lambda_i^\star > 0 \Rightarrow f_i(x^\star) = 0, \; f_i(x^\star) < 0 \Rightarrow \lambda_i^\star = 0
$$

Karush-Kuhn-Tucker (KKT) conditions

the following four conditions are called KKT conditions (for a problem with differentiable f_i, h_i):

1. primal constraints: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $h_i(x) = 0$, $i = 1, \ldots, p$
2. dual constraints: $\lambda \geq 0$
3. complementary slackness: $\lambda_i f_i(x) = 0$, $i = 1, \ldots, m$
4. gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

if strong duality holds and x, λ, ν are optimal, then they must satisfy the KKT conditions

KKT conditions for convex problem

if $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ satisfy KKT for a convex problem, then they are optimal:

- from complementary slackness: $f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- from 4th condition (and convexity): $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

hence, $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$.

if Slater's condition is satisfied, then x is optimal if and only if there exist λ, ν that satisfy KKT conditions

- recall that Slater's condition implies strong duality, and dual optimum is attained
- generalizes optimality condition $\nabla f_0(x) = 0$ for unconstrained problem

Optimality conditions for convex problems

- unconstrained convex problem: x is optimal if and only if
  $x \in \text{dom } f_0, \nabla f_0(x) = 0$
- equality constrained convex problem
  minimize $f_0(x)$ subject to $Ax = b$
  x is optimal if and only if there exists a ν such that
  $x \in \text{dom } f_0, Ax = b, \nabla f_0(x) + A^T\nu = 0$
- minimization over nonnegative orthant
  minimize $f_0(x)$ subject to $x \geq 0$
  x is optimal if and only if

$$x \in \mathrm{dom}\ f_0, x \geq 0, \begin{cases} \nabla f_0(x)_i \geq 0, & \text{if } x_i = 0 \\ \nabla f_0(x)_i = 0, & \text{if } x_i > 0, \end{cases} \quad \forall i$$

Outline

Quadratic programming problems

- Quadratic programming problem

$$\begin{aligned} \min_{x} \ & \frac{1}{2}x'Qx + c'x \\ \text{s.t. } & Ax = b \\ & x \geq 0, \end{aligned} \tag{1}$$

where $x \in \mathrm{R}^n$, $Q \in \mathrm{R}^{n\times n}$, $c \in \mathrm{R}^n$, $A \in \mathrm{R}^{m\times n}$, $Q \in \mathrm{R}^{n\times n}$, $b \in \mathrm{R}^m$

- Q is a symmetric matrix; if not, Q can be replaced by $(Q + Q')/2$
- The problem is convex if and only if Q is positive semidefinite.
- Mean-variance problem is a convex quadratic programming problem.

Optimality conditions for convex quadratic programs

Suppose Q is positive semidefinite. Then, x is an optimal solution to (1) if and only if there exists Lagrangian multiplier (s, y) such that the following four KKT conditions hold:

1. primal constraints: $Ax = b$, $x_i \geq 0, i = 1, \ldots, n$
2. dual constraints: $s \geq 0$
3. complementary slackness: $s_i x_i = 0$, $i = 1, \ldots, n$
4. gradient of Lagrangian with respect to x vanishes:

$$Qx + c - s + A'y = 0$$

Interior-point methods (IMP)

- Convex quadratic programming problems can be solved in polynomial time using IPMs
- The optimality KKT conditions in matrix form:

$$F(x,y,s) = \begin{bmatrix} Qx + c - s + A'y \\ Ax - b \\ XS\mathbf{1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, (x,s) \geq 0. \tag{2}$$

  Above, X and S are diagonal matrices with the entries of the x and s vectors, respectively, on the diagonal, i.e., $X_{ii} = x_i$, and $X_{ij} = 0, i \neq j$, and similarly for S.
- The system of equations $F(x,y,s) = 0$ is a square system. So we can apply Newton's method if we discard the nonnegativity constraints.
- IMP is based on a carefully modified Newton's method