IEOR E4007: Optimization Models and Methods

Quadratic Programming

Garud Iyengar
Columbia University
Industrial Engineering and Operations Research

Quadratic program

Optimization problem with

- Quadratic objective function
- Linear constraints
- Continuous variables

$$
\begin{array}{ll}
\text{min/max} & \underbrace{\sum_{i=1}^{d}\sum_{j=1}^{d} Q_{ij}x_i x_j}_{\text{quadratic function}} + \sum_{i=1}^{d} c_i x_i = x^\top Q x + c^\top x \\
\text{subject to} & Ax = b \\
 & Hx \geq g
\end{array}
$$

QP convex problem if, and only if, $f(x) = x^\top Q x + c^\top x$ is an appropriately convex/concave function.

- min (resp. max) $\Rightarrow$ $f(x)$ convex (resp. concave)

Example

$$
\begin{aligned}
f(x) &= 2x_1 + 3x_2 + 7x_1^2 + 3x_1x_2 + x_2^2 \\
&= \begin{bmatrix} 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^\top \underbrace{\begin{bmatrix} 7 & 3 \\ 0 & 1 \end{bmatrix}}_{Q \text{ is not symmetric}} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}
\end{aligned}
$$

$x^\top Qx = (x^\top Qx)^\top = x^\top Q^\top x = x^\top(\frac{1}{2}Q + \frac{1}{2}Q^\top)x$. Can symmetrize Q. Why bother?

$$
\begin{aligned}
f(x) &= 2x_1 + 3x_2 + \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^\top \left(\frac{1}{2} \begin{bmatrix} 7 & 3 \\ 0 & 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 7 & 0 \\ 3 & 1 \end{bmatrix} \right) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&= 2x_1 + 3x_2 + \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^\top \begin{bmatrix} 7 & 1.5 \\ 1.5 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}
\end{aligned}
$$

Assets and positions

Random total rate of returns on d assets: $\tilde{r} = (\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_d)^\top$

- Mean rate of return on asset j: $\mu_j = \mathbb{E}[\tilde{r}_j]$
- Covariance of returns on assets i and j: $V_{ij} = \mathbb{E}[(\tilde{r}_i - \mu_i)(\tilde{r}_j - \mu_j)]$

$$
V = \begin{bmatrix} V_{11} & V_{12} & \ldots & V_{1d} \\ \vdots & \vdots & \ddots & \vdots \\ V_{d1} & \ldots & \ldots & V_{dd} \end{bmatrix} = \mathbb{E}(\tilde{r} - \mu)(\tilde{r} - \mu)^\top
$$

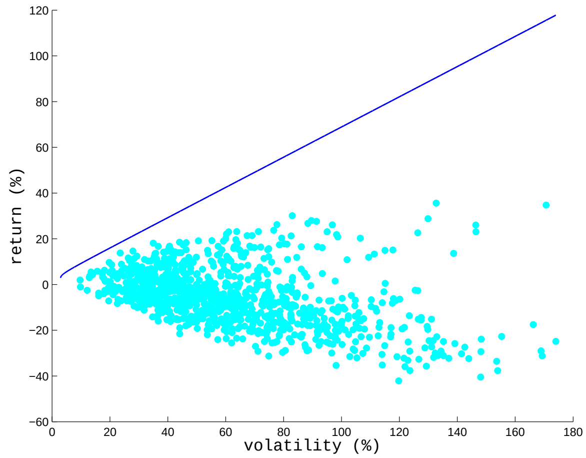
Dollar invested in the d assets: $w = (w_1, \ldots, w_d)^\top$

- Random return on position w: $\tilde{r}_w = \sum_{j=1}^d \tilde{r}_j w_j = \tilde{r}^\top w$
- Mean return on w: $\mu_w = \sum_{j=1}^d \mu_j w_j$
- Variance of return on w:

$$
\begin{aligned}
\sigma_w^2 &= \mathbb{E}(\tilde{r}_w - \mu_w)^2 \\
&= \mathbb{E}\big((\tilde{r} - \mu)^\top w\big)^2 \\
&= w^\top \mathbb{E}\big((\tilde{r} - \mu)(\tilde{r} - \mu)^\top\big) w = w^\top V w
\end{aligned}
$$

Portfolio selection problem

Efficient frontier



Two objectives:

- Maximize mean return μ_w
- Minimize volatility σ_w

Cannot do both ... have to construct a trade-off.

Efficient frontier: max return for a given volatility.

Several different formulations:

- Maximize return for a given volatility: $\max\{\mu_w : \sigma_w \leq \bar{\sigma}, w \in \mathcal{S}\}$
- Minimize volatility for a given return: $\min\{\sigma_w : \mu_w \geq \bar{\mu}, w \in \mathcal{S}\}$
- Maximize risk aversion adjusted return: $\max\{\mu_w - \tau\sigma_w^2 : w \in \mathcal{S}\}$

Time series analysis/prediction

Typical examples

- AR(m) process: $\tilde{r}_t = \sum_{k=1}^m \beta_k \tilde{r}_{t-k} + \epsilon_t$
- Factor model: $\tilde{r}_t = \sum_{k=1}^m \beta_k \tilde{f}_{t-1}^{(k)} + \epsilon_t$

Linear model

$$\tilde{y}_t = \beta^\top x_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \sigma^2).$$

Data: N observations

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}}_{y \in \mathbb{R}^N} = \underbrace{\begin{bmatrix} x_{11} & x_{21} & \ldots & x_{d1} \\ x_{12} & x_{22} & \ldots & x_{d2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1N} & x_{2N} & \ldots & x_{dN} \end{bmatrix}}_{X \in \mathbb{R}^{N \times d}} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \end{bmatrix}}_{\beta \in \mathbb{R}^d} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}}_{\epsilon \in \mathbb{R}^N}$$

$$y = X\beta + \epsilon$$

Likelihood of the data

$$\mathbb{P}(y \mid X) = \prod_{k=1}^N \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}|y_k - x_k^\top \beta|^2} = \frac{1}{(2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma^2}\|y - X\beta\|_2^2}$$

Modifications

Prior on β: Black-Litterman model

$$\min_{\beta} \|y - X\beta\|_2^2 + \lambda \|P(\beta - \beta_0)\|_2^2$$

Sparse estimate (LASSO):

$$\min_{\beta} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

The ℓ_1 norm $\|\beta\|_1 = \sum_{j=1}^{d} |\beta_j|$ is a sparsifying norm that induces a large number of components to be zero. Why do you think this happens? Why do we care?

Optimality conditions: Unconstrained problems

1D problem: $\min_{x\in\mathbb{R}}\{cx + qx^2\}$

$$\begin{aligned}
x^* \in \operatorname*{argmin}_{x\in\mathbb{R}} f(x) \quad &\Leftrightarrow \quad f(x^* + y) \geq f(x^*) \quad \forall y \\
&\Leftrightarrow \quad f(x^*) + (c + 2qx^*)y + qy^2 - f(x^*) \geq 0, \quad \forall y \\
&\Leftrightarrow \quad f'(x^*)y + qy^2 \geq 0, \quad \forall y \\
&\Leftrightarrow \quad f'(x^*) = 0, \quad q \geq 0.
\end{aligned}$$

What happens if $q < 0$? What is $\min_{x\in\mathbb{R}}\{cx + qx^2\}$?

f twice differentiable. f convex if, and only if, $f''(x) \geq 0$ for all x.

f quadratic. Then $\min_x f(x)$ is finite only if f is convex.

f quadratic. Then $\max_x f(x)$ is finite only if f is concave.

Optimality conditions: Unconstrained problems

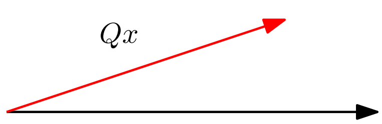
d-dimensional problem: $\min_{x\in\mathbb{R}^d} f(x) = \min_{x\in\mathbb{R}^d}\{c^\top x + x^\top Qx\}$

$$
\begin{aligned}
x^* \in \operatorname*{argmin}_{x\in\mathbb{R}^d} f(x) \quad &\Leftrightarrow \quad f(x^*+y) \geq f(x^*), \quad \forall y \in \mathbb{R}^d \\
&\Leftrightarrow \quad f(x^*) + (c + 2Qx^*)^\top y + y^\top Qy - f(x^*) \geq 0, \ \forall y \\
&\Leftrightarrow \quad \nabla f(x^*)^\top y + y^\top Qy \geq 0, \quad \forall y \in \mathbb{R}^d \\
&\Leftrightarrow \quad \nabla f(x^*) = 0, \quad y^\top Qy \geq 0, \quad \forall y \in \mathbb{R}^d
\end{aligned}
$$

The gradient of the function g is defined as

$$
\nabla g(x) = \begin{bmatrix} \frac{\partial g}{\partial x_1} \\ \frac{\partial g}{\partial x_2} \\ \vdots \\ \frac{\partial g}{\partial x_d} \end{bmatrix} \qquad \nabla f = \begin{bmatrix} c_1 + 2\sum_{j=1}^d Q_{1j}x_j \\ c_2 + 2\sum_{j=1}^d Q_{2j}x_j \\ \vdots \\ c_d + 2\sum_{j=1}^d Q_{dj}x_j \end{bmatrix} = c + 2Qx
$$

Eigenvalues and convexity



Q symmetric $d \times d$ matrix

- There exist d eigenvectors $\{v^{(1)}, \ldots, v^{(d)}\} \in \mathbb{R}^d$
- Eigenvectors are mutually orthonormal: $(v^{(i)})^\top v^{(j)} = 0$ if $i \neq j$ and 1 otherwise.
- $Q = \sum_{j=1}^d \lambda_j v^{(j)}(v^{(j)})^\top$

Using the eigenvalue-eigenvector expansion

$$
y^\top Qy = y^\top \Big(\sum_{j=1}^d \lambda_j v^{(j)}(v^{(j)})^\top \Big) y = \sum_{j=1}^d \lambda_j (y^\top v^{(j)})^2
$$

Thus

$$
y^\top Qy \geq 0 \ \forall y \quad \Leftrightarrow \quad \lambda_j \geq 0 \ \forall j \quad \Leftrightarrow \quad Q \text{ positive semidefinite}
$$

Optimality conditions: Unconstrained problems

$$x^* \in \underset{x\in\mathbb{R}^d}{\operatorname{argmin}} f(x) \quad \Leftrightarrow \quad \nabla f(x^*) = c + 2Qx^* = 0,\ Q \text{ positive semidefinite}$$

What if Q has a negative eigenvalue? Suppose $\lambda_1 < 0$. Then

$$f(x + \alpha v^{(1)}) = \alpha \nabla f(x)^\top v^{(1)} + \alpha^2 \lambda_1 \to -\infty$$

Maximization

$$x^* \in \underset{x\in\mathbb{R}^d}{\operatorname{argmax}} f(x) \quad \Leftrightarrow \quad \nabla f(x^*) = c + 2Qx^* = 0,\ Q \text{ negative semidefinite}$$

Simple example

Quadratic function of two variables

$$f(x) = 2x_1 + 3x_2 + \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^\top \underbrace{\begin{bmatrix} 7 & 1.5 \\ 1.5 & 1 \end{bmatrix}}_{Q} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Stationary points $\{x : \nabla f(x) = 0\}$. The function is stationary to first order at stationary points:

$$f(x + \delta x) \approx f(x) + \nabla f(x)^\top \delta x = f(x)$$

$$0 = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 14 & 3 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.2632 \\ -1.8947 \end{bmatrix}$$

maximum or minimum or neither? maximum or minimum or neither?

$$\texttt{eig}(Q) = \{0.6459, 7.3541\} \Rightarrow Q \text{ positive definite}$$

Simple example (contd)

Quadratic function of two variables

$$f(x) = 2x_1 + 3x_2 + \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^\top \underbrace{\begin{bmatrix} 7 & 1.5 \\ 1.5 & -1 \end{bmatrix}}_{Q} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Stationary point

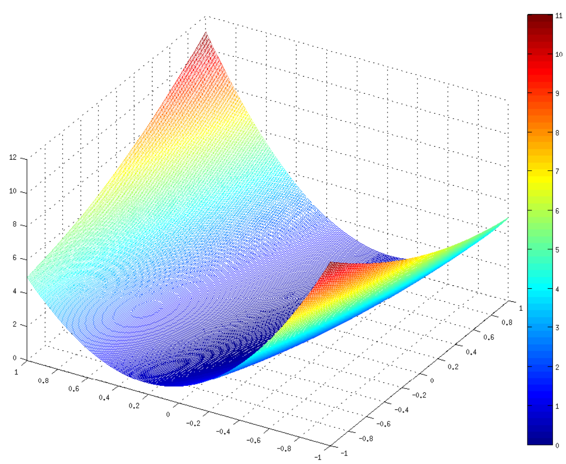
$$2Qx + c = 0 \quad \Rightarrow \quad x = -\frac{1}{2}Q^{-1}c = \begin{bmatrix} -0.3514 \\ 0.9730 \end{bmatrix}$$

Eigenvalues of Q: $\{-1.2720, 7.2720\}$

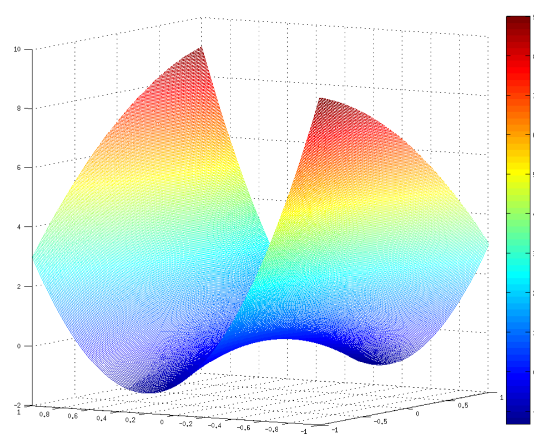
Maximum, minimum, or neither? Maximum, minimum, or neither.

Simple example plots

Positive definite Q



Indefinite Q



“cup” shaped function

“saddle” shaped function

Least squares problem

Optimization problem

$$\min_{\beta \in \mathbb{R}^d} \|y - X\beta\|_2^2 \equiv \min_{\beta \in \mathbb{R}^d} \left\{ \beta^\top (X^\top X)\beta - 2(X^\top y)^\top \beta \right\}$$

Is $X^\top X$ positive semidefinite? $y^\top X^\top X y = \|Xy\|_2^2 \geq 0$. Yes.

Stationary point: $2(X^\top X)\beta - 2(X^\top y) = 0 \Rightarrow \beta = (X^\top X)^{-1}(X^\top y)$

Problems?

- What is X is not full rank?
- What if X is full rank but has very small eigenvalues?

Numerical stability

Finite precision and round-off errors

- Accumulate over operations
- Can be very serious for high-dimensional problems
- Equivalent operations can have very different impact on errors

Inverse is numerically unstable. Use QR decomposition.

$$[Q, R] = \texttt{qr}(A) \quad \Rightarrow \quad A = QR, \quad Q \text{ unitary}, \quad R \text{ upper triangular}$$

$$Ax = b \quad \Leftrightarrow \quad Rx = Q^\top b \text{ and now back substitute}$$

Note: No inverses!

Punch line: In MATLAB use $x = A \backslash b$ and **NOT** $x = \texttt{inv}(A) * b$

Optimality conditions: constrained problems

Generic optimization problem $\min\{f(x) : x \in \mathcal{S}\}$

- x^* global optimum if

$$f(x^*) \leq f(y) \text{ for all } y \in \mathcal{S}$$

- x^* local optimum if there exists $\delta_x > 0$ such that

$$f(x^*) \leq f(y) \text{ for all } y \in \mathcal{S} \cap \{z : \|z - x\|_2 \leq \delta_x\}$$

  x^* is optimal in local neighborhood.

Want to compute global optima but have to settle for local optima!

w is a feasible direction at x if there exists $\alpha_0 > 0$ such that

$$x + \alpha w \in \mathcal{S} \text{ for all } 0 \leq \alpha \leq \alpha_0.$$

One can move a positive amount in direction w and still remain feasible.

Constrained QPs: Feasible directions

z feasible for the constrained QP

$$\begin{array}{rl} \text{min} & c^\top x + x^\top Q x \\ \text{s.t.} & Ax = b \\ & Hx \geq g \end{array}$$

Let $\mathcal{F} = \{x : Ax = b, Hx \geq g\}$

Feasible directions at z: $z + \alpha w \in \mathcal{F}$ for all small enough α

$$W = \left\{ w : \begin{array}{l} Aw = 0 \\ H_z w \geq 0 \end{array} \right\}$$

H_z = matrix corresponding to active constraints at z

z stationary or critical point if $\nabla f(z)^\top w \geq 0$ for all $w \in W$

Local minima $\Rightarrow$ critical point (not the other way!)

- Equality constraints: $Aw = 0$
- Inactive inequality constraint: satisfied for all small enough α

$$h_i^\top (x + \alpha w) \geq g_i \quad \Leftrightarrow \quad \alpha h_i^\top w \geq \underbrace{g_i - h_i^\top x}$$

Characterizing critical points

$$
\begin{aligned}
z \text{ critical point} \quad &\Leftrightarrow \quad \nabla f(z)^\top w \geq 0, \text{ for all } Aw = 0,\ H_z w \geq 0 \\
&\Leftrightarrow \quad \min\{\nabla f(z)^\top w : Aw = 0,\ H_z w \geq 0\} = 0
\end{aligned}
$$

Linear programming duality: strong duality holds since primal feasible

$$
\begin{array}{llcll}
\text{min} & \nabla f(z)^\top w & = & \text{max} & 0^\top u + 0^\top v \\
\text{s. t.} & Aw = 0 & & \text{s. t.} & A^\top u + H_z^\top v = \nabla f(z) \\
 & H_z w \geq 0 & & & v \geq 0
\end{array}
$$

Equivalent representation

- Gradient condition: $\sum_{i=1}^{m_{eq}} a_i u_i + \sum_{i=1}^{m_{ineq}} h_i v_i = 2Qz + c$
- Sign constraints on duals: $v_i \geq 0$
- Complementary slackness: $v_i(h_i^\top z - g_i) = 0$

Problem: Cannot use these conditions to compute a critical point. Only check whether a point is critical.

Convex problem: critical points are global optima

Convex optimization problem: $\min\{f(x) : x \in \mathcal{S}\}$

Feasible directions at $z \in \mathcal{S}$: $W = \mathcal{S} - z$

$$y \in \mathcal{S} \Rightarrow z + \theta(y - z) \in \mathcal{S} \text{ for all } \theta \in [0, 1]$$

Facts

- f convex: $f(y) \geq f(x) + \nabla f(x)^\top (y - x)$ for all x, y
- z critical point and $W = \mathcal{S} - z$: $\nabla f(z)^\top (y - z) \geq 0$ for all $y \in \mathcal{S}$

Conclusion: $f(y) \geq f(z) + \nabla f(z)^\top (y - z) \geq f(z)$

z critical point for convex opt problem $\Rightarrow$ z global opt

Only equality constraints

No sign constraints on duals and no complementary slackness.

- Primal feasibility: $Ax = b$ (m_{eq} equations)
- Dual feasibility: $c + 2Qx = A^\top u$ (d equations)
- $m_{eq} + d$ unknowns: $x \in \mathbb{R}^d$, $u \in \mathbb{R}^{m_{eq}}$

$$\begin{bmatrix} A & 0 \\ 2Q & -A^\top \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} = \begin{bmatrix} b \\ -c \end{bmatrix}$$

Typically, solved in stages:

- Solve for x in terms of u: $x = \underbrace{\frac{1}{2} Q \backslash A^\top}_{P} u - \underbrace{\frac{1}{2} Q \backslash c}_{s}$
- Solve for u: $Ax = APu - As = b$, i.e. $u = (AP) \backslash (b + As)$
- Compute $x = Pu - s$.

Simple mean-variance portfolio selection

Minimum variance problem

$$\begin{array}{ll} \text{min} & x^\top V x \\ \text{s.t.} & \mu^\top x = r \\ & \mathbf{1}^\top x = 1 \end{array}$$

Note:

- $\mu^\top x = r$ instead of $\mu^\top x \geq r$
- V positive semidefinite: critical points are optimal

Calculation for computing optimal solution

$$2Vx = \begin{bmatrix} \mu & \mathbf{1} \end{bmatrix} u \quad \Rightarrow \quad x = \tfrac{1}{2} \begin{bmatrix} V^{-1}\mu & V^{-1}\mathbf{1} \end{bmatrix} u$$

$$\underbrace{\tfrac{1}{2} \begin{bmatrix} \mu^\top \\ \mathbf{1}^\top \end{bmatrix} V^{-1} \begin{bmatrix} \mu & \mathbf{1} \end{bmatrix}}_{M^{-1} \in \mathbb{R}^2} u = \begin{bmatrix} r \\ 1 \end{bmatrix} \quad \Rightarrow \quad u = 2M \begin{bmatrix} r \\ 1 \end{bmatrix} = \begin{bmatrix} M_{11} r + M_{12} \\ M_{21} r + M_{22} \end{bmatrix}$$

$$\begin{aligned} x^* = \tfrac{1}{2} \begin{bmatrix} V^{-1}\mu & V^{-1}\mathbf{1} \end{bmatrix} u \quad &= \quad \Big(M_{11} V^{-1}\mu + M_{12} V^{-1}\mathbf{1} \Big) r \\ & \quad\quad + \Big(M_{12} V^{-1}\mu + M_{22} V^{-1}\mathbf{1} \Big) \end{aligned}$$

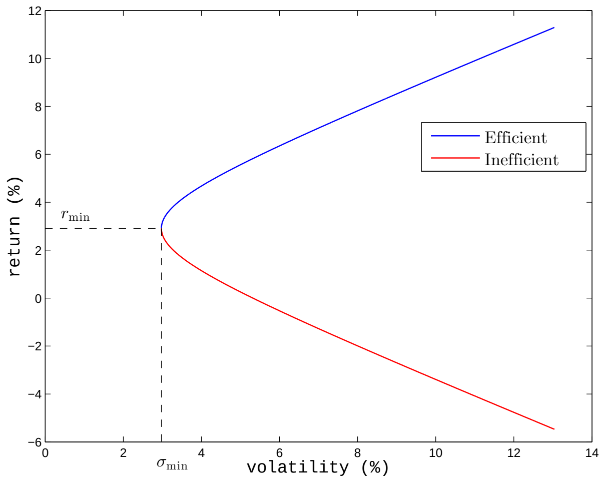
Simple mean-variance portfolio selection

The optimal portfolio has a linear structure

$$x^*(r) = g + hr, \qquad g, h \in \mathbb{R}^d$$

Optimal variance

$$\sigma^2(r) = x^*(r)^\top V x^*(r) = (g^\top V g)r^2 + 2(g^\top V h)r + h^\top V h$$



Why did we get the inefficient part?

Because we insisted on equality in the mean constraint.

Two-fund theorem

Consider two different return values $r_1 \neq r_2$. Then

$$\left.\begin{array}{rcl} x^*(r_1) & = & g + hr_1 \\ x^*(r_2) & = & g + hr_2 \end{array}\right\} \Rightarrow \begin{array}{rcl} g & = & \frac{1}{r_2 - r_1}\Big(r_2 x^*(r_1) - r_1 x^*(r_2)\Big) \\ h & = & \frac{1}{r_2 - r_1}\Big(x^*(r_2) - x^*(r_1)\Big) \end{array}$$

Therefore,

$$x^*(r) = g + hr = \left(\frac{r_2 - r}{r_2 - r_1}\right) x^*(r_1) + \left(\frac{r - r_1}{r_2 - r_1}\right) x^*(r_2)$$

So, why is this important?

Theorem. (Two-fund theorem) The mean-variance optimal portfolio for any return r can be constructed by diversifying over two mutual funds.

This result ultimately leads to the Capital Asset Pricing Model (CAPM)

Inequality constrained QPs

Optimization problem

$$\begin{array}{ll} \min & x^\top Q x + c^\top x \\ \text{s.t.} & Ax = b \\ & Hx \geq g \end{array} \qquad \begin{array}{ll} \min & x^\top V x \\ \text{s.t.} & \mu^\top x = R \\ & \mathbf{1}^\top x = 1 \\ & x \geq -K\mathbf{1} \end{array}$$

Only know how to compute critical points for equality constrained QPs

- Solve a sequence of equality constrained QPs
- Ensure that a critical point for the equality constrained problem is also a critical for the full QP

Step 1: Set $t = 0$. Compute a feasible point $x^{(t)}$. How?

Let I denote the indices of the inequality constraints active at $x^{(t)}$.

Step 2: Create an equality constrained QP by setting

$$Ax = b, \qquad h_i^\top x = g_i, \quad i \in I.$$

Is this a relaxation or restriction of the original QP?

Active set algorithm: step

Equality constrained QP

$$\begin{array}{ll} \min & x^\top Q x + c^\top x \\ \text{s.t.} & Ax = b, \quad H_I x = g_I \end{array}$$

Portfolio example Suppose $x_j^{(t)} = -K$ for all $j = j_1, \ldots, j_s$. Then

$$I = \{j_1, \ldots, j_s\}.$$

If $x_j = -K$, keep it fixed there.

More convenient to write $x = x^{(t)} + y$: Compute critical point for

$$\begin{array}{ll} \min & y^\top Q y + \nabla f(x^{(t)})^\top y \\ \text{s.t.} & Ay = 0, \quad H_I y = 0 \end{array}$$

Let y^* denote the critical point, and let (u, v_I) denote the duals.

Step 3: Suppose $y^* \neq 0$. Compute maximum step length $\lambda \leq 1$ such that $x^{(t)} + \lambda y^*$ is feasible.

$$\lambda^* = \max\left\{\lambda : h_i^\top (x^{(t)} + \lambda y^*) \geq g_i, \forall i \notin I, \lambda \leq 1\right\}$$

Why do we want $\lambda^* \leq 1$?

Active set algorithm: updating the active set

Step 3 (contd): Set $x^{(t+1)} = x^{(t)} + \lambda^* y^*$. Set $t \leftarrow t+1$. Go to **Step 2**.

Step 4: Suppose $y^* = 0$, i.e. $x^{(t)}$ is a critical point for its active set of constraints. Need to check if $x^{(t)}$ is a critical point for the original QP.

$x^{(t)}$ is primal feasible. Have to ensure duals have the correct sign.

- duals corresponding to equality constraints: u
- duals corresponding to inactive inequality constraints: set $v_{I^c} = 0$
- duals corresponding to the active inequality constraint: $v_I \geq 0$

Note that in the equality constrained QP the dual v_I are free.

Suppose v_I are all non-negative. Then $x^{(t)}$ is the required critical point.

Suppose $v_i < 0$ for some $i \in I$. Since $v_i = \frac{\partial \text{obj}}{\partial g_i} < 0$ it follows that objective will improve by allowing the constraint to go slack or inactive. Set $I = I \backslash \{i\}$. Go to **Step 2**.

Portfolio selection example

4 asset example in MATLAB file `activeport.m` for $K = 1.0$

Iteration 0

$$
\begin{aligned}
x &= \begin{bmatrix} 0.2886 & 0 & 0.7114 & 0 \end{bmatrix} \\
y &= \begin{bmatrix} 0.1395 & -0.0014 & 1.7250 & -1.8631 \end{bmatrix} \\
u &= \begin{bmatrix} 0.1489 & -0.1565 \end{bmatrix} \\
v &= \begin{bmatrix} 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix}
\end{aligned}
$$

Iteration 1

$$
\begin{aligned}
x &= \begin{bmatrix} 0.3635 & -0.0008 & 1.6373 & -1.0000 \end{bmatrix} \\
y &= \begin{bmatrix} -0.0028 & 0.0354 & -0.0326 & 0.0000 \end{bmatrix} \\
u &= \begin{bmatrix} 0.2092 & -0.2212 \end{bmatrix} \\
v &= \begin{bmatrix} 0.0000 & 0.0000 & 0.0000 & 0.0015 \end{bmatrix}
\end{aligned}
$$

Portfolio selection example

Iteration 2

$$
\begin{aligned}
x &= \begin{bmatrix} 0.3607 & 0.0346 & 1.6047 & -1.0000 \end{bmatrix} \\
y &= \begin{bmatrix} 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix} \\
u &= \begin{bmatrix} 0.2092 & -0.2212 \end{bmatrix} \\
v &= \begin{bmatrix} 0.0000 & 0.0000 & 0.0000 & 0.0015 \end{bmatrix}
\end{aligned}
$$

Sharpe ratio optimization

Optimization problem

$$
\begin{aligned}
P = \quad \text{sup} \quad & \tfrac{\mu^\top x}{\sqrt{x^\top S x}} \\
\text{s.t.} \quad & \mathbf{1}^\top x = 1 \\
& Ax \leq b
\end{aligned}
$$

where $S \succeq 0$ and $Ax \leq b$ are additional side constraints on the problem. Assume there exists a feasible portfolio $\bar{x}$ such that $\mu^\top \bar{x} > 0$.

Not a quadratic program or a linear program. But, the objective is **invariant** under positive scaling.

$$
f(x) = \frac{\mu^\top x}{\sqrt{x^\top S x}} \quad \Rightarrow \quad f(x) = f(\alpha x) \text{ for all } \alpha > 0
$$

Will use this invariance to solve the Sharpe ratio problem by solving a related QP.

Sharpe ratio optimization (contd)

Consider the following QP in (y, α):

$$\begin{array}{lll} \inf & y^\top S y & \\ \text{s.t.} & \mu^\top y = 1, & \mathbf{1}^\top y = \alpha \\ & Ay \le \alpha b, & \alpha \ge 0 \end{array}$$

Clearly $(\bar{y} = \frac{1}{\bar{\mu}^\top \bar{x}}\bar{x}, \bar{\alpha} = \frac{1}{\mu^\top \bar{x}})$ is feasible for this QP.

By adding the redundant constraint $y^\top S y \le \bar{y}^\top S \bar{y}$, we can conclude that an optimal solution (y^*, α^*) exists.

Note that this QP is equivalent to the optimization problem

$$\begin{array}{llll} Q = & \max & \frac{1}{y^\top S y} & \\ & \text{s.t.} & \mu^\top y = 1, & \mathbf{1}^\top y = \alpha \\ & & Ay \le \alpha b, & \alpha \ge 0 \end{array}$$

The optimal solution of this problem is also (y^*, α^*)!

Sharpe ratio optimization (contd)

Claim: $P = Q$

Proof: Suppose x is feasible for P with $\mu^\top x > 0$. Then $(y = \frac{1}{\mu^\top x}x, \frac{1}{\mu^\top x})$ is feasible for Q with the same objective value. Therefore, $Q \ge P$.

Next we consider two cases:

(1) $\alpha^* > 0$: $x = \frac{1}{\alpha^*}y^*$ is feasible for P with an the same objective value. Therefore, $P \ge Q$. In this case, the optimal solution for P is $x^* = \frac{1}{\alpha^*}y^*$.

(2) $\alpha^* = 0$: $x = \bar{x} + \gamma y^*$ is feasible for all $\gamma > 0$. Moreover,

$$P \ge \lim_{\gamma \to \infty} \frac{\mu^\top x}{\sqrt{x^\top S x}} = \frac{1}{\sqrt{(y^*)^\top S y^*}} = Q$$

So, it follows that $P = Q$. But the value P is achieved with infinite leverage, i.e. $\gamma = \infty$.

Second-order cone programs

Extension of quadratic program with much more modeling power.

Second-order cone (SOC) constraint $\|Ax+b\|_2 \le c^\top x + d$. Example

$$\left\| \begin{bmatrix} 2 & 3 \\ 6 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} -1 \\ 4 \end{bmatrix} \right\|_2 \le \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 12$$

Convex quadratic constriants are SOC constraints: $Q \succeq \mathbf{0} \Rightarrow Q = V^\top V$ for some $V \in \mathbb{R}^{r \times d}$

$$\begin{aligned} x^\top Q x \le \sigma^2 \quad &\Leftrightarrow \quad x^\top V^\top V x \le \sigma^2 \\ &\Leftrightarrow \quad \|Vx\|_2 \le \sigma \end{aligned}$$

$$\begin{aligned} x^\top Q x \le 4z \quad &\Leftrightarrow \quad x^\top V^\top V x \le (1+z)^2 - (1-z)^2 \\ &\Leftrightarrow \quad x^\top V^\top V x + (1-z)^2 \le (1+z)^2 \\ &\Leftrightarrow \quad \left\| \begin{bmatrix} Vx \\ 1-z \end{bmatrix} \right\|_2 \le (1+z) \end{aligned}$$

Why did we just not take square roots on both side?

SOC portfolio selection problems

mean-volatility portfolio selection problem

$$\begin{array}{ll} \max & \mu^\top x - \tau\sqrt{x^\top Q x} \\ \text{s.t.} & x \in \mathcal{S} \end{array} \quad \equiv \quad \begin{array}{ll} \max & \mu^\top x - \tau z \\ \text{s.t.} & \|Vx\|_2 \le z \\ & x \in \mathcal{S} \end{array}$$

mean-volatility portfolio selection with multiple scenarios

$$\begin{array}{ll} \max & \min_{1 \le k \le N} \left\{ \mu_k^\top x - \tau\sqrt{x^\top Q_k x} \right\} \\ \text{s.t.} & x \in \mathcal{S} \end{array} \quad \equiv \quad \begin{array}{ll} \max & z \\ \text{s.t.} & \mu_k^\top x - \tau \|V_k x\| \ge z \\ & x \in \mathcal{S} \end{array}$$

robust mean-variance portfolio selection: $\mu \in \mathcal{M} = \{\mu_0 + Pu : \|u\|_2 \le 1\}$

$$\min_{\mu \in \mathcal{M}} \{\mu^\top x\} = \mu_0^\top x + \min_{\|u\|_2 \le 1} \{u^\top P^\top x\} = \mu_0^\top x - \|P^\top x\|_2$$

$$\begin{array}{ll} \max & \min_{\mu \in \mathcal{M}} \{\mu^\top x\} - \tau x^\top Q x \\ \text{s.t.} & x \in \mathcal{S} \end{array} \quad \equiv \quad \begin{array}{ll} \max & z - 4\tau y \\ \text{s.t.} & \mu_0^\top x - \|P^\top x\|_2 \ge z \\ & \left\| \begin{bmatrix} Vx \\ 1-y \end{bmatrix} \right\|_2 \le (1+y) \end{array}$$

General nonlinear programming

$$\begin{array}{rll} \text{min} & f(x) & \\ \text{subject to} & a_i(x) = b_i, & i = 1, \ldots, m, \\ & h_j(x) \geq g_j, & j = 1, \ldots, p, \\ & x \in \mathbb{R}^d & \end{array}$$

$$\left.\begin{array}{r} f(x) \text{ convex} \\ a_i(x) \text{ linear} \\ h_j(x) \text{ concave} \end{array}\right\} \quad \begin{array}{c} \Rightarrow \\ \not\Leftarrow \end{array} \quad \text{convex optimization problem}$$

Optimality conditions for unconstrained problems:

- $\bar{x}$ local optimal $\Rightarrow \nabla f(\bar{x}) = \mathbf{0}$, $\nabla^2 f(\bar{x}) \succeq \mathbf{0}$ (positive semidefinite)
- $\nabla f(\bar{x}) = \mathbf{0}$, $\nabla^2 f(\bar{x}) \succ \mathbf{0}$ (positive definite) $\Rightarrow \bar{x}$ local optimal

Hessian $\nabla^2 f(x)$ is not a constant but a function of x. $\nabla^2 f(\bar{x}) \succeq \mathbf{0}$ is not sufficient only necessary.

Examples of unconstrained nonlinear problems

One dimensional example

- $f(x) = x^2$: $\nabla f(x) = 2x = 0 \Rightarrow x = 0$, $\nabla^2 f(0) = 2 > 0$ local min
- $f(x) = x^3$: $\nabla f(x) = 3x^2 = 0 \Rightarrow x = 0$, $\nabla^2 f(0) = 6x \mid_{x=0} = 0$ possibly local min (saddle pt.)
- $f(x) = x^4$: $\nabla f(x) = 4x^3 = 0 \Rightarrow x = 0$, $\nabla^2 f(0) = 12x^2 \mid_{x=0} = 0$ possibly local min (local min)

2 dimensional example

$$f(x) = 3x_1^2 x_2 + 6x_1^3 + 2x_2^2$$

$$\nabla f(x) = \begin{bmatrix} 6x_1x_2 + 18x_1^2 \\ 3x_1^2 + 4x_2 \end{bmatrix} \quad \nabla^2 f(x) = \begin{bmatrix} 6x_2 + 36x_1 & 6x_1 \\ 6x_1 & 4 \end{bmatrix}$$

Critical points are hard to compute in closed form.

Computing critical points: gradient based method

Current iterate: x_k

Gradient at current iterate: $g_k = \boldsymbol{\nabla} f(x_k)$

Move in the direction $-\boldsymbol{\nabla} f(x_k)$ direction. Optimal step

$$t^* = \operatorname*{argmin}_{t \geq 0} f\big(x_k + t\boldsymbol{\nabla} f(x_k)\big)$$

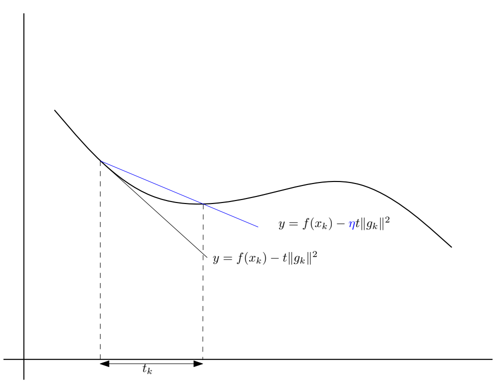
$h(t) = f\big(x_k + t\boldsymbol{\nabla} f(x_k)\big)$ is a 1-dimensional function of t. Setting the derivative equal to zero, we get

$$0 = h'(t^*) = \boldsymbol{\nabla} f\big(x_k + t^*\boldsymbol{\nabla} f(x_k)\big)^\top \boldsymbol{\nabla} f(x_k) = \boldsymbol{\nabla} f(x_{k+1})^\top \boldsymbol{\nabla} f(x_k)$$

Search direction at iterate x_{k+1} is orthogonal to the search direction at x_k

- Gradient method zig-zags and is very slow
- Conjugate-gradient method fixes this

Line search in gradient method



Current iterate: x_k
Gradient at current iterate: $g_k = \boldsymbol{\nabla} f(x_k)$
How to select step t: $x = x_k - tg_k$

"Expected" function value

$$f(x) \approx f(x_k) - t\,\|g_k\|^2$$

Fix $0 < \eta < 1$. Choose the largest step t such that

$$f(x - tg_k) \leq f(x_k) - \eta t\,\|g_k\|^2$$

Make at least η fraction of the progress predicted by the gradient.

Stochastic gradient method

Let $\mathcal{D} = \{(x_i, y_i) : 1 \leq i \leq N\}$ denote a data set with $N \gg 1$

Let $\ell(x_i, y_i; \theta)$ denote a "risk" function, e.g.

$$\ell(x_i, y_i; \theta) = (y_i - (\alpha + \beta^\top x_i))^2, \qquad \theta = (\alpha, \beta)$$

Empirical Risk Minimization problem:

$$\min_{\theta} \ f(\theta) = \frac{1}{N} \sum_{i=1}^{N} \ell(x_i, y_i; \theta)$$

Computing $\boldsymbol{\nabla} f(\theta) = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\nabla} \ell(x_i, y_i; \theta)$ is expensive since $N \gg 1$

- Uniformly sample a data point $i \in \{1, \ldots, N\}$
- Use $g(\theta) = \boldsymbol{\nabla} \ell(x_i, y_i; \theta)$ instead of $\boldsymbol{\nabla} f(\theta)$
- Note that $\boldsymbol{\nabla} f(\theta) = \mathbb{E}[g(\theta)]$

Can use importance sampling and control variates to improve performance

Conjugate gradient method

Modification of the gradient method that avoids the zig-zag behavior

Choose search direction p_k, $k \geq 0$ as follows:

- If $k = 0$, i.e. we are at initial point. Set $p_0 = -\boldsymbol{\nabla} f(x_0)$
- Else set

$$p_k = -\boldsymbol{\nabla} f(x_k) + \beta_k p_{k-1} \qquad \beta_k = \frac{\|\boldsymbol{\nabla} f(x_k)\|_2^2}{\|\boldsymbol{\nabla} f(x_{k-1})\|_2^2}$$

Next iterate: $x_{k+1} = x_k + t_k p_k$ where t_k is the step length

The update is a gradient step modified by a momentum term

$$x_{k+1} = x_k - t_k \boldsymbol{\nabla} f(x_k) - \underbrace{\frac{t_k \beta_k}{t_{k-1}} (x_k - x_{k-1})}_{\text{momentum}}$$

Computing critical points: Hessian based method

Minimize the quadratic approximation

$$\hat{f}(x) = f(x_k) + \nabla f(x_k)^\top (x - x_k) + (x - x_k)^\top \nabla^2 f(x_k)(x - x_k)$$

Then

$$x_{k+1} = \operatorname*{argmin}_{x \in \mathbb{R}^d} \hat{f}(x) = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k)$$

But, is $\delta = x_{k+1} - x_k$ a descent direction, i.e.

$$\nabla f(x_k)^\top \delta = -\nabla f(x_k)^\top \nabla^2 f(x_k)^{-1} \nabla f(x_k) \leq 0$$

Not always! Guaranteed only if $\nabla^2 f(x_k)$ positive semidefinite.

Variable metric / quasi-Newton method

Iteratively approximate $\nabla^2 f(x_k)^{-1}$ by a positive definite matrix B_k

Let $y_{k+1} = g_{k+1} - g_k$ and $s_{k+1} = x_{k+1} - x_k$. Want

$$\begin{aligned} s_{k+1} &= B_{k+1} y_{k+1} \\ &= (B_k + \Delta B_k) y_{k+1} \end{aligned}$$

Rank-2 update $\Delta B_k = a s_{k+1} s_{k+1}^\top + b(B_k y_{k+1})(B_k y_{k+1})^\top$. Thus,

$$s_{k+1} = B_k y_{k+1} + a(s_{k+1}^\top y_{k+1}) s_{k+1} + b(y_k^\top B_k y_{k+1})(B_k y_{k+1})$$

One solution: $a = \frac{1}{s_{k+1}^\top y_{k+1}}$ and $b = -\frac{1}{y_{k+1}^\top B_k y_{k+1}}$. Thus

$$B_{k+1} = B_k + \frac{s_{k+1} s_{k+1}^\top}{s_{k+1}^\top y_{k+1}} - \frac{B_k y_{k+1} y_{k+1}^\top B_k}{y_{k+1}^\top B_k y_{k+1}}$$

Known as the Davidon-Fletcher-Powell method

Variable metric / quasi-Newton method

Iteratively approximate $\nabla^2 f(x_k)$ by a positive definite matrix H_k

Interchanges the role of $y_{k+1} = g_{k+1} - g_k$ and $s_{k+1} = x_{k+1} - x_k$

$$H_{k+1} = H_k + \frac{s_{y+1} s_{y+1}^\top}{s_{k+1}^\top y_{k+1}} - \frac{H_k s_{k+1} s_{k+1}^\top H_k}{s_{k+1}^\top H_k s_{k+1}}$$

Thus,

$$H_{k+1}^{-1} = \left(\mathbf{I} - \frac{s_{k+1} y_{k+1}^\top}{s_{k+1}^\top y_{k+1}} \right) H_{k-1}^{-1} \left(\mathbf{I} - \frac{s_{k+1} y_{k+1}^\top}{s_{k+1}^\top y_{k+1}} \right) + \frac{s_{k+1} s_{k+1}^\top}{y_{k+1}^\top s_{k+1}}$$

Known as the Broyden-Fletcher-Goldfarb-Shanno method

- More often used in practice
- A limited memory version of this method is very popular

Critical points: constrained problems

Approximate all non-linear functions by a first-order Taylors' series expansion

$$\begin{array}{rll} \min & f(x) & \\ \text{s.t.} & a_i(x) = b_i, & \forall i, \\ & h_j(x) \geq g_j, & \forall j, \\ & x \in \mathbb{R}^d & \end{array} \quad \Rightarrow \quad \begin{array}{rll} \min & f(\bar{x}) + \nabla f(\bar{x})^\top (x - \bar{x}) & \\ \text{s.t.} & a_i(\bar{x}) + \nabla a_i(\bar{x})^\top (x - \bar{x}) = b_i, & \forall i, \\ & h_j(\bar{x}) + \nabla h_j(\bar{x})^\top (x - \bar{x}) \geq g_j, & \forall j, \\ & x \in \mathbb{R}^d & \end{array}$$

The $\bar{x}$ is a local minimum and a constraint qualification such as

- gradients of all active constraints are linearly independent
- all constraints are linear

is satisfied. Then $\bar{x}$ is a critical point of the linearized problem, i.e. there exists dual multipliers u and v such that

- Gradient condition: $\nabla f(\bar{x}) = \sum_{i=1}^{m} u_i \nabla a_i(\bar{x}) + \sum_{j=1}^{p} v_j \nabla h_j(\bar{x})$
- Sign constraints on duals: $v_j \geq 0$
- Complementary slackness: $v_j(h_j(\bar{x}) - g_j) = 0$

Karush-Kuhn-Tucker point

A point $\bar{x}$ that is a critical point of the linearized problem is also called the Karush-Kuhn-Tucker (KKT) point.

The KKT point for a convex optimization problem is globally optimal.

How does one compute a KKT point?

- Compute a feasible descent direction d by solving the LP

$$\begin{array}{rll} \text{min} & \nabla f(x_k)^\top d & \\ \text{s.t.} & \nabla a_i(x_k)^\top d = 0, & \forall i, \\ & \nabla h_j(x_k)^\top d \geq 0, & \forall j : h_j(x_k) = b_j \text{ (active)} \end{array}$$

- If $d = 0$ we are at a KKT point.
- If not, do a line search along d to obtain the next iterate x_{k+1}

KKT example

Is $(x_1, x_2, x_3) = (0, 2, 1)$ a KKT point for the optimization problem?

$$\begin{array}{rl} \max & f(x) \equiv \ln(3x_1 + 2x_2) - 5x_3 \\ \text{s.t.} & g_1(x) \equiv 3x_1 + 2x_2 \geq 1 \\ & g_2(x) \equiv 4x_2 - x_3 \leq 7 \end{array}$$

- Tight constraints?

$$3x_1 + 2x_2 = 2 > 1 \qquad 4x_2 - x_3 = 7 \leq 7$$

- Gradient condition

$$\nabla f(x) = \begin{bmatrix} \frac{3}{3x_1+2x_2} \\ \frac{2}{3x_1+2x_2} \\ -5 \end{bmatrix} = \begin{bmatrix} \frac{3}{4} \\ \frac{2}{4} \\ -5 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \\ -1 \end{bmatrix} u, \qquad u \geq 0$$

Optimization direction

Compute an optimizing direction d that keeps all the current active constraints active

$$\begin{array}{rl} \text{max} & \nabla f(x)^\top d \\ \text{subject to} & \nabla g_j(x)^\top d = 0, \quad \forall j = 1, \ldots, m \\ & \|d\|_2 \le 1 \end{array}$$

Dual problem

$$\text{min}_\lambda \Big\| \nabla f(x) - \sum_{j=1}^{m} \nabla g_j(x)\lambda_j \Big\|_2$$

Optimal d is the projection of $\nabla f(x)$ on the tangent space, or equivalently, the residual remaining after projecting on the normals.
Back to the example:

$$d = \nabla f(x) - (\nabla f(x)^\top \nabla g_2(x)) \frac{\nabla g_2(x)}{\|\nabla g_2(x)\|}$$

KKT Example

Consider the optimization problem

$$\begin{array}{rl} \text{max} & c^\top x \\ \text{s.t.} & g_1(x) \equiv \|x + a\|_2 \le \|a\|_2 \\ & g_2(x) \equiv \|a - x\|_2 \le \|a\|_2 \end{array}$$

Only feasible point $x = 0$ ... local min

- $\nabla g_1(0) = \frac{x+a}{\|x+a\|_2}|_{x=0} = \frac{a}{\|a\|_2}$
- $\nabla g_2(0) = -\frac{a-x}{\|a-x\|_2}|_{x=0} = -\frac{a}{\|a\|_2}$

Gradient condition

$$c = \frac{a}{\|a\|_2}(u_1 - u_2), \quad u_1, u_2 \ge 0$$

No solution if $c \neq ua$.