

6. 限界は破れるか（1）

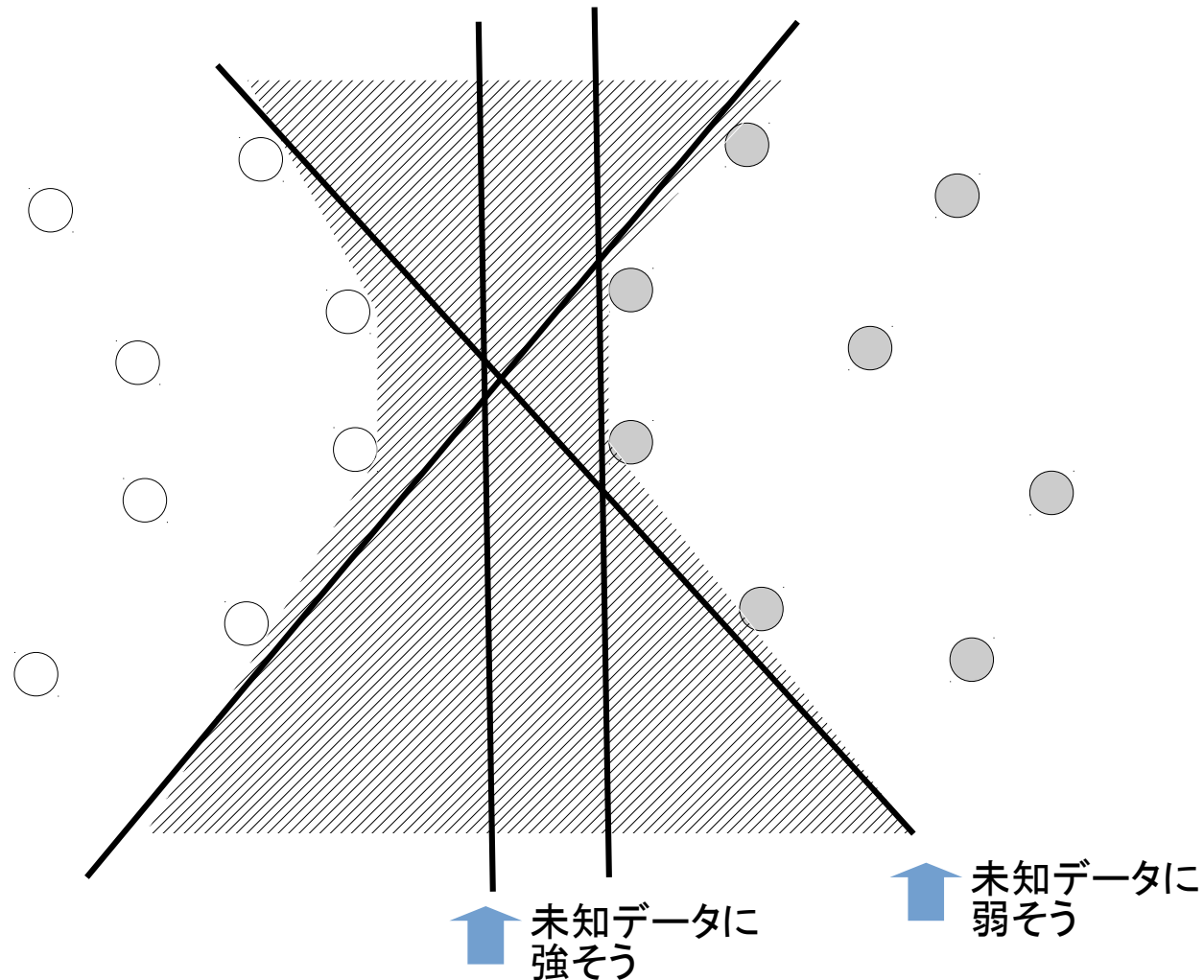
ー サポートベクトルマシン ー

- パーセプトロンの学習規則の限界
 - 学習パターンが線形分離可能である場合は識別面が見つかるが、信頼できる識別面とは限らない
 - 学習パターンが線形分離不可能である場合は、学習が停止しない

 サポートベクトルマシン（SVM）

6.1 識別面は見つかったけれど

パーセプトロンの学習規則ではどれが見つかるかわからない

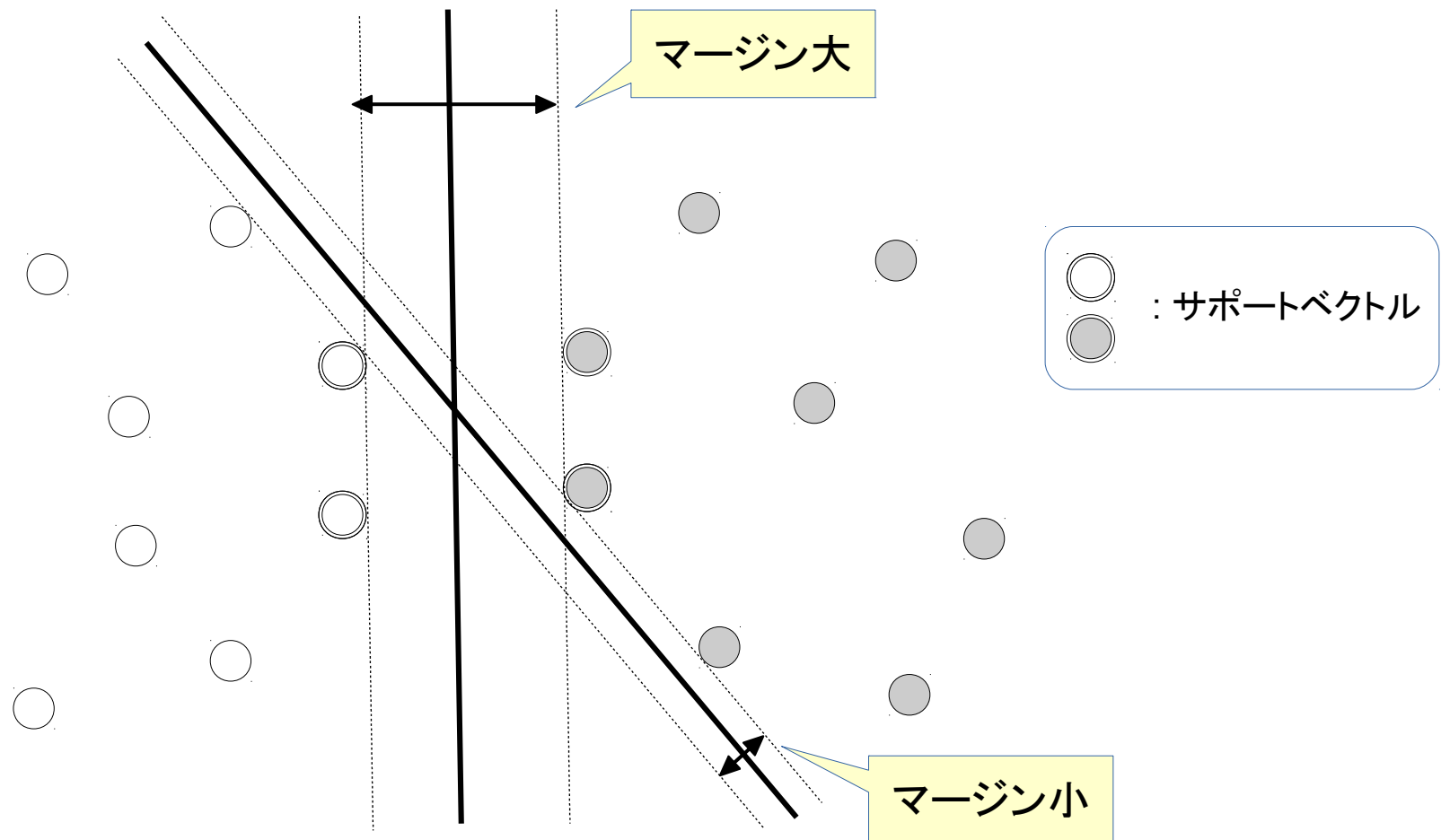


6.2 サポートベクトルマシンの学習アルゴリズム

6.2.1 サポートベクトル

- 線形SVM

- マージン最大となる線形識別面を求める



6.2.2 マージンを最大にする

- 学習データ

$$\{(\mathbf{x}_i, y_i)\} \quad i = 1, \dots, n, \quad y_i = 1 \text{ or } -1$$

- 線形識別面の式

$$\mathbf{w}^T \mathbf{x} + w_0 = 0$$

- 識別面の制約の導入（係数を定数倍しても平面は不変）

$$\min_{i=1, \dots, n} |\mathbf{w}^T \mathbf{x}_i + w_0| = 1$$

- 学習パターンと識別面との最小距離（＝マージン）

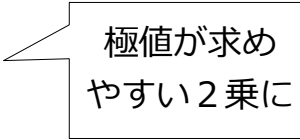
$$\min_{i=1, \dots, n} \text{Dist}(\mathbf{x}_i) = \min_{i=1, \dots, n} \frac{|\mathbf{w}^T \mathbf{x}_i + w_0|}{\|\mathbf{w}\|} = \frac{1}{\|\mathbf{w}\|}$$

これを
最大化

点と直線の距離の公式

$$r = \frac{|ax + by + c|}{\sqrt{a^2 + b^2}}$$

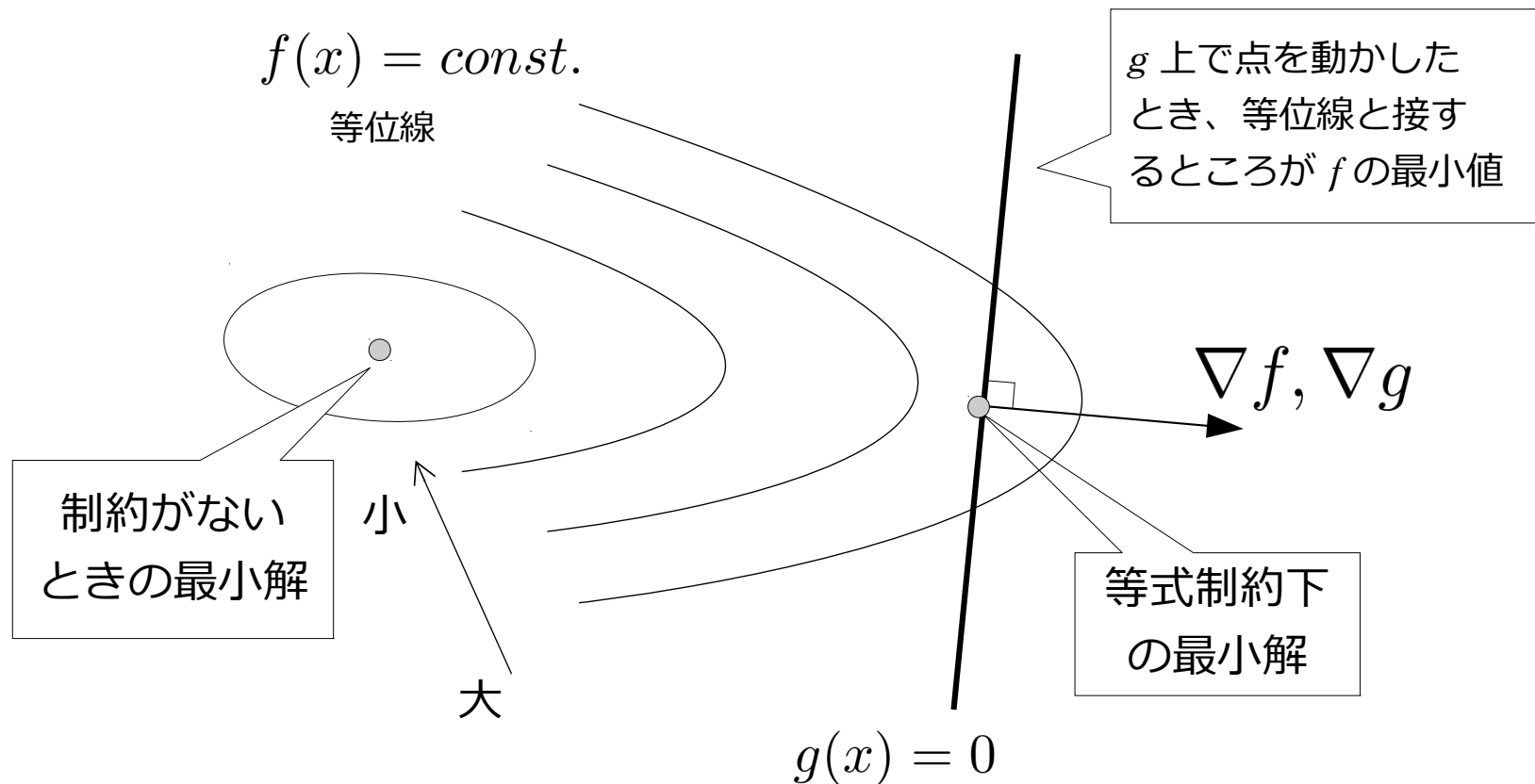
6.2.2 マージンを最大にする

- 目的関数の置き換え： $\min \frac{1}{2} \|\mathbf{w}\|^2$  極値が求めやすい2乗に
- 制約条件： $y_i(\mathbf{w}^T \mathbf{x}_i + w_0) \geq 1 \quad i = 1, \dots, n$
- 解法：ラグランジュの未定乗数法
 - 問題 $\min f(x) \quad s.t. \quad g(x) = 0$
 - ラグランジュ関数 $L(x, \alpha) = f(x) - \alpha g(x)$
 - $\alpha \geq 0$
 - x, α で偏微分して0になる値が極値

ラグランジュの未定乗数法 (付録A.4)

$$\min f(x) \quad s.t. \quad g(x) = 0 \quad \frac{\partial L(x, \alpha)}{\partial x} = \nabla f(x) - \alpha \nabla g(x) = 0$$

$$L(x, \alpha) = f(x) - \alpha g(x) \quad \frac{\partial L(x, \alpha)}{\partial \alpha} = -g(x) = 0$$



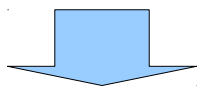
6.2.2 マージンを最大にする

- 計算

$$L(\mathbf{w}, w_0, \boldsymbol{\alpha}) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i (y_i (\mathbf{w}^T \mathbf{x} + w_0) - 1)$$

$$\frac{\partial L}{\partial w_0} = 0 \quad \Rightarrow \quad \sum_{i=1}^n \alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \quad \Rightarrow \quad \mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i$$



$$L(\boldsymbol{\alpha}) = \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j - \sum_{i=1}^n \alpha_i$$

$$\alpha_i \geq 0$$

最大化2次計画問題
→ $\boldsymbol{\alpha}$ が求まる

6.2.2 マージンを最大にする

- 定数項の計算
 - 各クラスのサポートベクトルから求める

$$w_0 = -\frac{1}{2}(\boldsymbol{w}^T \boldsymbol{x}_{s1} + \boldsymbol{w}^T \boldsymbol{x}_{s2})$$

- 識別関数

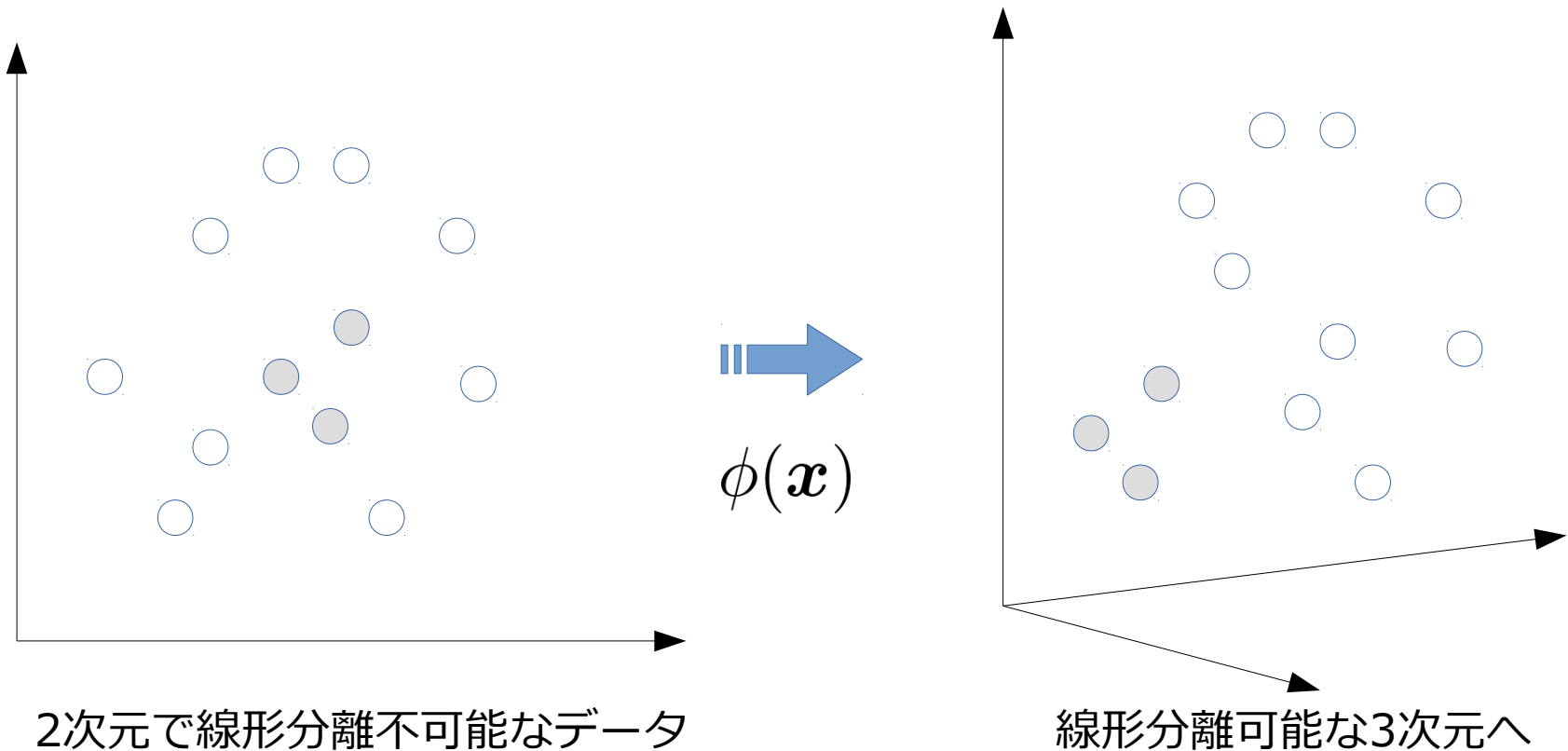
$$\begin{aligned} g(\boldsymbol{x}) &= \boldsymbol{w}^T \boldsymbol{x} + w_0 \\ &= \sum_{i=1}^n \alpha_i y_i \boldsymbol{x}^T \boldsymbol{x}_i + w_0 \end{aligned}$$

サポートベクトルに対応する α_i のみが0以上、残りは0

6.3 線形分離可能にしてしまう

6.3.1 高次元空間への写像

- 特徴ベクトルの次元数を増やす



ただし、元の空間でのデータ間の
距離関係は保持するように

6.3.2 カーネル法

- 非線形変換関数： $\phi(\mathbf{x})$
- カーネル関数
 - 元の空間での距離が変換後の空間の内積に対応

$$K(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{x}')$$

線形カーネル
 $(\mathbf{x}^T \mathbf{x}')^p$
を用いる場合もある

- カーネル関数の例

– 多項式カーネル $K(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}' + 1)^p$

– ガウシアンカーネル $K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{\sigma^2}\right)$

これらの形であれば、対応する非線形変換が存在することが数学的に保証されている

6.3.2 カーネル法

- 変換後の識別関数： $g(\boldsymbol{x}) = \boldsymbol{w}^T \phi(\boldsymbol{x}) + w_0$
- SVMで求めた \boldsymbol{w} の値を代入

$$\begin{aligned} g(\boldsymbol{x}) &= \sum_{i=1}^n \alpha_i y_i \phi(\boldsymbol{x})^T \phi(\boldsymbol{x}_i) + w_0 \\ &= \sum_{i=1}^n \alpha_i y_i K(\boldsymbol{x}, \boldsymbol{x}_i) + w_0 \end{aligned}$$

非線形変換の
式は不要！！！！

カーネルトリック

6.3.3 具体的なカーネル関数

- 線形カーネル（2次）の展開

$$\begin{aligned}K(\mathbf{x}, \mathbf{x}') &= (\mathbf{x}^T \mathbf{x}')^2 \\&= (x_1 x'_1 + x_2 x'_2)^2 \\&= x_1^2 x'^2_1 + 2x_1 x_2 x'_1 x'_2 + x_2^2 x'^2_2 \\&= (x_1^2, \sqrt{2}x_1 x_2, x_2^2) \cdot (x'^2_1, \sqrt{2}x'_1 x'_2, x'^2_2)\end{aligned}$$

- 多項式カーネル（2次）の展開

$$\begin{aligned}K(\mathbf{x}, \mathbf{x}') &= (\mathbf{x}^T \mathbf{x}' + 1)^2 \\&= (x_1 x'_1 + x_2 x'_2 + 1)^2 \\&= x_1^2 x'^2_1 + x_2^2 x'^2_2 + 2x_1 x_2 x'_1 x'_2 + 2x_1 x'_1 + 2x_2 x'_2 + 1 \\&= (x_1^2, x_2^2, \sqrt{2}x_1 x_2, \sqrt{2}x_1, \sqrt{2}x_2, 1) \cdot \\&\quad (x'^2_1, x'^2_2, \sqrt{2}x'_1 x'_2, \sqrt{2}x'_1, \sqrt{2}x'_2, 1)\end{aligned}$$